Investigation of Asparagine-Linked Glycosylation in Archaeal and Bacterial Systems

by

Angelyn Larkin

B.A. Chemistry and Political Science
Wellesley College, 2003

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
FEB 11 2011
LIBRARIES
ARCHIVES

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

December 2010
[February 2011]

© 2010 Massachusetts Institute of Technology
All rights reserved

Signature of Auth _____
Department of Chemistry
December 10, 2010

Certified by: _____
Barbara Imperiali
Class of 1922 Professor of Chemistry and Professor of Biology
Thesis Supervisor

Accepted by: _____
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Professor Sarah E. O'Connor: _____

Chair
Associate Professor of Chemistry

Professor Barbara Imperiali: _____

Thesis Supervisor
Class of 1922 Professor of Chemistry and Professor of Biology

Professor Catherine L. Drennan: _____

Professor of Chemistry and Biology
Investigator and Professor, Howard Hughes Medical Institute

Investigation of Asparagine-Linked Glycosylation in Archaeal and Bacterial Systems

by

Angelyn Larkin

Submitted to the Department of Chemistry
on December 10, 2010 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ABSTRACT

Asparagine-linked protein glycosylation entails the stepwise assembly of an oligosaccharide onto a polyisoprenyl diphosphate carrier, followed by the *en bloc* transfer of the glycan onto acceptor proteins by oligosaccharyl transferase (OTase). It is now clear that this *N*-linked protein modification catalyzed by OTases is conserved in all three kingdoms of life. In contrast to eukaryotic OTases, which are multimeric complexes made up of several membrane-spanning proteins, bacterial and archaeal OTases instead appear to be composed of just a single membrane-bound subunit, offering a more tractable system for detailed biochemical characterization. Although significant progress has been made with the bacterial OTase PglB from *Campylobacter jejuni*, problems with low protein expression yields and poor stability have complicated in depth study. In order to identify a more suitable OTase candidate, a selection of archaeal OTases was screened for heterologous expression levels in *Escherichia coli*, and it was determined that the homolog from the methanogen *Methanococcus voltae* possessed the best protein expression profile, indicating a 100-fold improvement over PglB.

In an effort to generate a substrate to probe the function of the *M. voltae* OTase, we required a robust synthesis of the highly modified UDP-GlcNAc(3NAc)A and thus turned to the Wbp pathway from the opportunistic pathogen *Pseudomonas aeruginosa*. Early genetic studies suggested that *P. aeruginosa* produces UDP-GlcNAc(3NAc)A as a precursor to ManNAc(3NAc)A, a carbohydrate found in the O-antigen of the lipopolysaccharide of the organism. Using a combination of protein biochemistry and NMR spectroscopy, three enzymes (WbpB, WbpE, and WbpD) were confirmed to be responsible for the biosynthesis of UDP-GlcNAc(3NAc)A. It is shown that WbpB and WbpE are a dehydrogenase/aminotransferase pair that converts UDP-GlcNAcA to UDP-GlcNAc($3NH_2$)A in a coupled reaction via a unique NAD^+ recycling pathway. In addition, the X-ray crystal structure of WbpE was solved in complex with its PLP cofactor and UDP-GlcNAc($3NH_2$)A product as the external aldimine. With UDP-GlcNAc(3NAc)A in hand, preliminary steps towards completion of the desired dolichyl-linked substrate for the study of the *M. voltae* OTase are described.

Finally, biochemical studies were undertaken in an attempt to inhibit the *C. jejuni* OTase, PglB. To this end, a panel of isosteric peptides was synthesized to identify possible PglB inhibitors. In addition, treatment of PglB with residue-specific alkylating agents coupled with site-directed mutagenesis revealed a key histidine residue that may play an important role in enzyme catalysis. Taken together, these studies offer insight into long-standing questions about the mechanism of oligosaccharide transfer.

Thesis Supervisor: Barbara Imperiali
Title: Class of 1922 Professor of Chemistry and Professor of Biology

Acknowledgments

I would first like to extend my deepest gratitude to my advisor, Barbara Imperiali. I am inspired by the fearlessness and enthusiasm with which you have tackled some of the most difficult research problems. Thank you for your support and guidance over the years, and for fostering an enormously stimulating and collaborative working environment. It has been an honor to be a part of your group. I would also like to thank the members of my thesis committee, Professors Sarah O'Connor and Cathy Drennan, for their thoughtful advice and encouragement throughout my time at MIT.

I am also indebted to the many creative and compassionate teachers I have had the privilege to learn from over the course my education, particularly Professor David Haines at Wellesley College. Thank you for encouraging my love of science and believing in me.

Over the past few years, I have been extremely fortunate to work with a fantastic group of labmates that have also become my friends. Elvedin, Matthieu, and Galen, thank you for warmly welcoming me into the lab and providing countless hours of sound advice and entertainment. Your friendship means a lot to me. Meredith and Brenda, I am grateful for your good cheer and constant encouragement, and am very glad that we have gone through this journey together. I will miss our daily chats. Nelson, thank you for sharing your love of protein crystallography with me and offering support in countless ways. Jay, a great podmate, I have enjoyed our talks about science and life. Mary and Eranthie, I often can't believe that we only overlapped in the lab for a short time; thank you for your advice and support over the years.

To the rest of the Imperiali lab, I am always amazed at my luck for ending up with such kind-hearted, collaborative, and fun coworkers. Thank you to Jebrell, Beth, Bianca, Seungjib, and Dora for getting me started. Anne, James, Wendy, Langdon, Mark, Marcie, Michelle, Mike, Andreas, Juan Antonio, Susana, Cliff, Andrew, Elke, Vinita, and Philipp, it's been a pleasure working and learning with you; thank you for all you've taught me and for making grad school so much fun. I am grateful to Elizabeth for all the she does to help the lab run smoothly. And to Marcie, Michelle, Mike, Vinita, and Philipp: I look forward to hearing about your future discoveries in the exciting yet often challenging field of *N*-linked glycosylation. Best of luck to you!

And of course, I am sure that I wouldn't have made it this far without the guidance, encouragement, and loving support of my amazing family. To Mom, Dad, Marya, and Tom, for being a constant source of strength, and for selflessly providing your compassion, loyalty, and humor, I dedicate this thesis to you. Thank you to Barb, Barry, and the Homrighaus family for your kindness and unwavering support. And most of all to James, my husband and best friend, I am grateful every day for your presence in my life. Thank you for everything.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

List of Abbreviations

Standard 3-letter and 1-letter codes are used for the 20 natural amino acids.
Standard 1-letter codes are used for the 4 common DNA bases.

AcCoA	acetyl-coenzyme A
amu	atomic mass unit
Bicine	*N*,*N*-Bis(2-hydroxyethyl)glycine
Bis-Tris	2-Bis-(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol
BSA	bovine serum albumin
CAM	ceric ammonium molybdate
CAZy	Carbohydrate-Active EnZYme database
CCP4	Collaborative Computational Project No. 4
CDI	1,1'carbonyldiimidazole
CE	capillary electrophoresis
CoA	coenzyme A
CHAPSO	3-[(3-cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate
CHES	2-(*N*-cyclohexylamino)ethanesulfonic acid
COSY	correlation spectroscopy
CPS	capsular polysaccharide
Dab	2,4-diaminobutanoic acid
DDM	n-dodecyl-β-D-maltoside
DEPC	diethylpyrocarbonate
DIPEA	diisopropylethylamine
DMA	dimethylacetamide
DMF	dimethylformamide
DMSO	dimethylsulfoxide
Dol-P	dolichylphosphate
Dol-PP-GlcNAc	dolichyldiphosphate *N*-acetylglucosamine
Dol-PP-GalNAc	dolichyldiphosphate *N*-acetylgalactosamine
dTDP	deoxy-thymidine 5'-phosphate
DTT	dithiothreitol
ε	extinction coefficient for molar absorptivity
EDT	ethanedithiol
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
FAD	flavin adenine dinucleotide
Fmoc	9-fluorenylmethoxycarbonyl
GDP-Man	guanosine diphosphate mannose
GPI	glycosylphosphoinositol
GST	glutathione S-transferase
IAA	iodoacetamide
IPTG	isopropyl β-D-1-thiogalactopyranoside

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMBC	heteronuclear multiple quantum coherence
HOBt	1-hydroxybenzotriazole
HSQC	heteronuclear single quantum coherence
2-HG	2-hydroxyglutarate
α-KG	α-ketoglutarate
K_{cat}	catalytic constant
K_i	inhibition constant
K_m	Michaelis constant
LB	Luria-Bertani broth
LDAO	lauryl dimethylamine oxide
LPS	lipopolysaccharide
MAD	multiwavelength anomalous diffraction
MALDI-TOF	matrix assisted laser desorption/ionization time of flight
MES	4-morpholineethanesulfonic acid
MIR	multiple isomorphous replacement
MMTS	methyl methanethiolsulfonate
MR	molecular replacement
MS	mass spectrometry
MWCO	molecular weight cutoff
NAD^+	nicotinamide adenine dinucleotide
Ni-NTA	nickel nitrilotriacetic acid
NMR	nuclear magnetic resonance
Nph	4-nitrophenylalanine
OPD	*o*-phenylenediamine
PAL	5-(4'aminomethyl-3',5'-dimethyoxyphenoxy)valeric acid
PDB	Protein Data Bank
PEG-PS	polyethyleneglycol-grafted polystyrene
PLP	pyridoxal 5'-phosphate
PMB	*para*-methoxybenzyl
PMP	pyridoxamine 5'-phosphate
PSUP	pure solvent upper phase
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
r.m.s.d	root mean square deviation
RP-HPLC	reversed-phase high-performance liquid chromatography
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeMet	selenomethionine
TEAB	triethylammonium bicarbonate
TEV	tobacco etch protease
THF	tetrahydrofuran
TFA	trifluoroacetic acid
TIS	triisopropylsilane
TLC	thin-layer chromatography
TNBS	2,4,6-trinitrobenzenesulfonic acid
TUP	theoretical upper phase

UDP	uridine 5'-diphosphate
UDP-GlcA	UDP-D-glucuronic acid
UDP-GlcNAc	UDP-*N*-acetyl-D-glucosamine
UDP-GlcNAcA	UDP-*N*-acetyl-D-glucosaminuronic acid
UDP-GlcNAc(3keto)A	UDP-2-acetamido-2-deoxy-3-oxo-D-glucuronic acid
UDP-GlcNAc(3NH_2)A	UDP-2-acetamido-3-amino-2,3-dideoxy-D-glucuronic acid
UDP-GlcNAc(3NAc)A	UDP-2,3-diacetamido-2,3-dideoxy-D-glucuronic acid
UDP-ManNAc(3NAc)A	UDP-2,3-diacetamido-2,3-dideoxy-D-mannuronic acid
Und-P	undecaprenylphosphate
UV	ultraviolet
Xaa	used to denote any amino acid
Xan	xanthenol

Chapter 1 Introduction

Introduction

Glycosylation is the most abundant protein modification found in nature, occurring across all kingdoms of life.[1] The great variety of carbohydrates coupled with the vast number of possible carbohydrate linkages serve to diversify the proteome in a manner beyond what is directly encoded in the genome. The appendage of a glycan to a protein can profoundly impact its function, playing important roles in a number of biological processes such as protein stability and rigidity, intracellular localization, cellular signaling and adhesion, and the immune response.

There are three major types of protein glycosylation that have been observed in nature (Figure 1-1): *N*-linked, *O*-linked, and GPI-anchored. *N*-linked glycosylation, the focus of this chapter, involves the *en bloc* transfer of an oligosaccharide onto the side chain amide nitrogen of asparagine residues within acceptor proteins. In the case of *O*-linked glycosylation, monosaccharides are typically ligated one at a time onto the side chain hydroxyl oxygen atom of either serine or threonine residues; however, in recent years, glycosylation of other residues such as tyrosine, hydroxylysine, and hydroxyproline has also been identified.[2,3] In the third type, a complex glycosylphosphoinositol moiety (GPI anchor) is transferred to the C-terminus of a target protein, serving to tether it to the cell membrane. Other classes of protein glycosylation have also been described in addition to these three main types, including phosphoglycosylation, in which sugar 1-phosphates are added to the hydroxyl oxygen atom of serine,[4,5] and *C*-mannosylation, where mannose residues are transferred to the C2' indole carbon of tryptophan residues,[6] although the biochemical details of these modifications are still unclear and require further investigation.

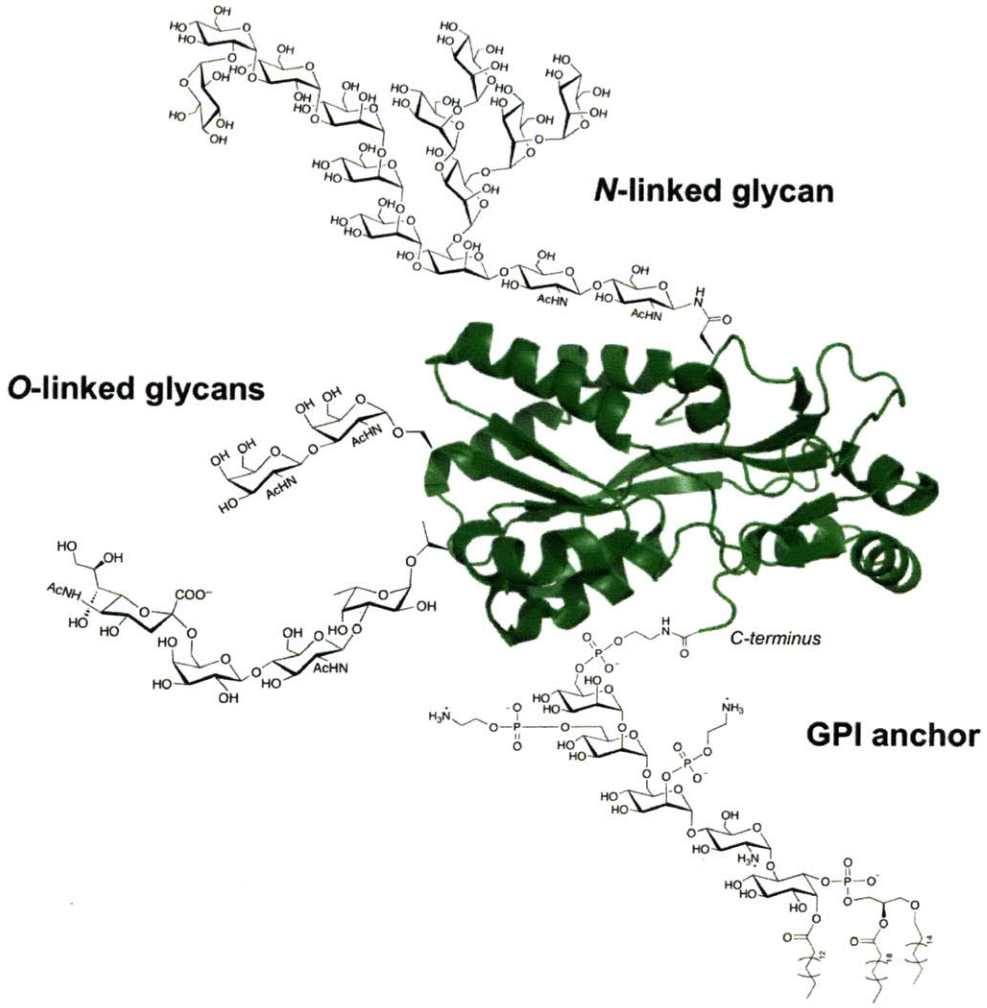


Figure 1-1: Three major classes of protein glycosylation.

While *N*-linked glycosylation was once believed to be limited only to eukaryotes, it is now firmly established that this complex modification occurs in both bacteria and archaea as well.[7,8] In the last ten years, the field of *N*-linked glycosylation has made enormous strides in elucidation of the enzymes involved in glycan assembly and processing, discovery of new and unusual carbohydrates, and the biological impact that these glycan modifications impart on the function of target proteins. In this chapter, the process of *N*-linked glycosylation will be

discussed across the three kingdoms of life. Special focus will be placed on the different means by which the glycans are assembled, the variety in the glycan structures, and how this structural diversity imparts function to specific glycoproteins. In addition, a brief overview of the role that *N*-linked glycosylation plays in viruses will be presented as well.

N-linked Glycosylation in Eukaryotes

Glycan Assembly

In eukaryotes, *N*-linked glycosylation occurs at the membrane of the endoplasmic reticulum (ER), and involves the transfer of a tetradecasaccharide ($Glc_3Man_9GlcNAc_2$) from a dolichyldiphosphate carrier onto the amide side chain nitrogen of an acceptor protein. Although a majority of the genetic and biochemical characterization of this pathway has been achieved in *Saccharomyces cerevisiae*, the process is remarkably conserved in all eukaryotes, from yeast to man.[9] Assembly of the glycan is carried out by a series of membrane-bound glycosyltransferases in the Alg family (*a*sparagine-*l*inked *g*lycosylation), which catalyze the transfer of each monosaccharide onto a dolichyldiphosphate carrier. Dolichol, a long chain α-saturated polyisoprene that varies in size between 14 and 21 units depending on the cell type and species, is phosphorylated by the kinase Sec59 prior to glycan assembly.[10,11] In addition to its role as a membrane anchor, dolichylphosphate may serve other functions in the glycosylation process as well, such promoting membrane fluidity, facilitating translocation across the ER membrane, and recruitment of key enzymes to the site of glycosylation.[12-14]

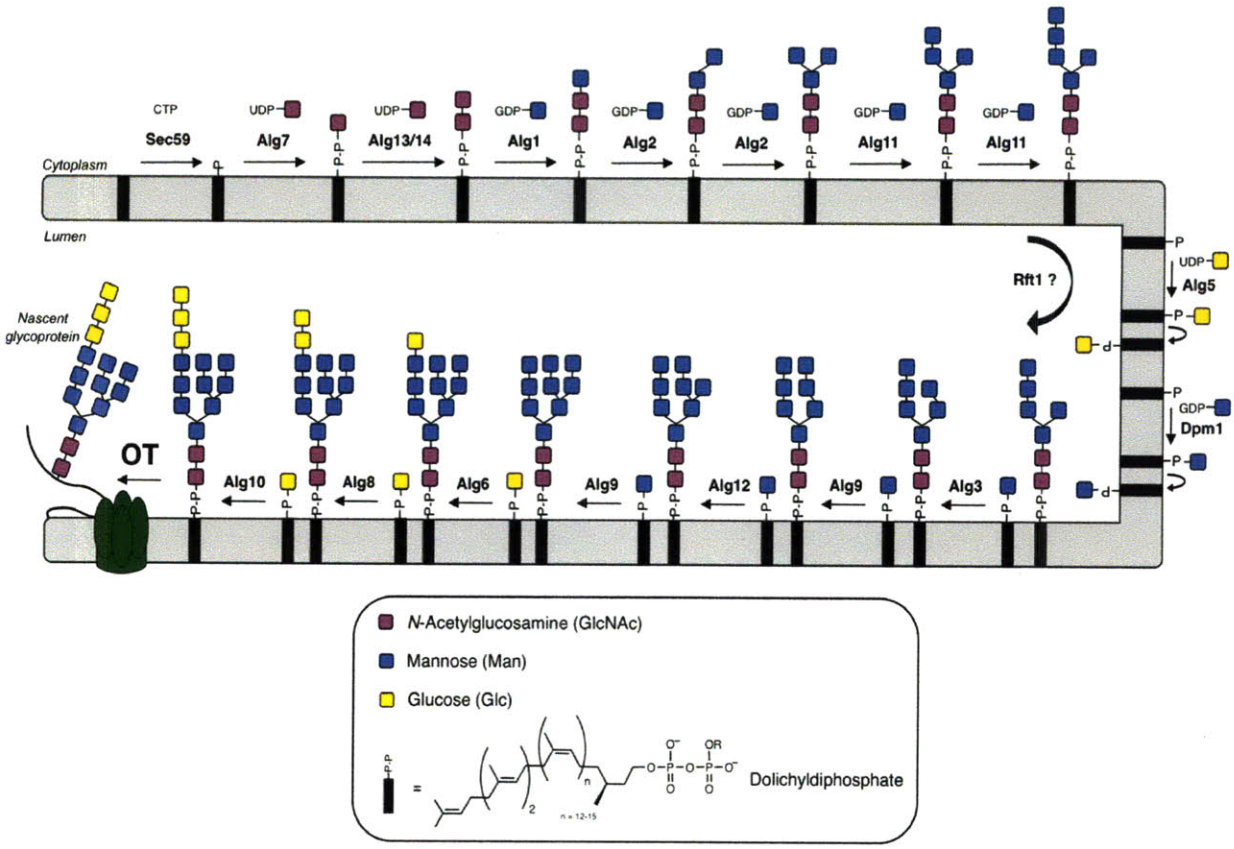


Figure 1-2: Pathway of *N*-linked glycosylation at the ER membrane in *S. cerevisiae*. In the first half of the process, the glycan is assembled from nucleotide sugar donors before being flipped to the ER lumen. The glycosyl donors in the lumen are Dol-P-Man and Dol-P-Glc, the biosynthesis of which is shown on the right.

The first phase of glycan assembly takes place on the cytoplasmic side of the ER membrane, where a high concentration of the nucleotide sugars donors (UDP-GlcNAc or GDP-Man) act as substrates for the early glycosyltransferases. The pathway begins with the addition of GlcNAc-1-P to the Dol-P carrier by Alg7,[15] followed by the transfer of the second GlcNAc residue by the Alg13/14 complex.[16,17] Recent work has established that the C-terminal domains of both Alg13 and Alg14 interact to form a functional glycosyltransferase capable of catalyzing GlcNAc transfer.[18,19] The first mannosylation is performed by Alg1 and serves as the primary branching point in the glycan structure.[20,21] Alg2 and Alg11 then sequentially install the next

four mannose residues, in which Alg2 carries out α-1,6 and α-1,3 mannosylations, followed by two iterative α-1,2 mannose additions by Alg11 to yield the heptasaccharide.[22-25] Genetic and biochemical studies have recently indicated that these early enzymes form discrete protein complexes in the ER membrane that may be important for maintaining flux through the pathway.[26,27] After completion of the heptasaccharide, the Dol-PP-$GlcNAc_2Man_5$ intermediate is flipped from the cytoplasmic to the lumenal face of the ER by a process that remains unclear. While early genetic studies implicated Rft1 as the required flippase,[28] recent biochemical evidence has suggested that this protein is in fact not required for translocation of the oligosaccharide intermediate, although the identity of the enzyme responsible has not yet been determined.[29-32] Further work in this area is necessary to characterize this elusive step.

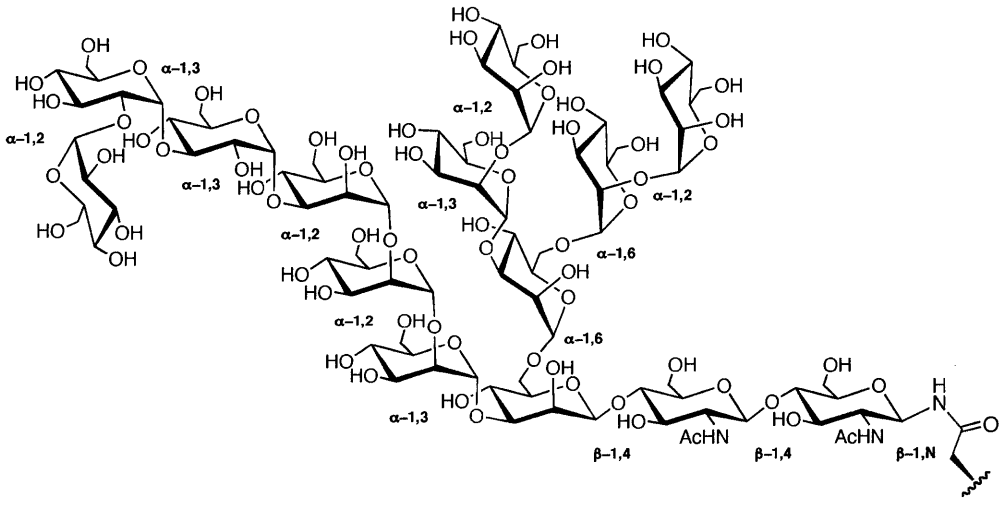


Figure 1-3: Structure of the eukaryotic *N*-linked glycan, $Glc_3Man_9GlcNAc_2$. The nature of the linkage between each carbohydrate is specified.

Glycan assembly continues on the lumenal side of the ER membrane, where the remaining four mannosylations are catalyzed by the actions of Alg3, Alg9, Alg12, and Alg9;[33]

Alg9 has been shown to transfer both α-1,2-linked mannose residues that cap two of the three branches of the glycan.[34] The three terminal glucose residues are then installed onto the nascent oligosaccharide by Alg6, Alg8, and Alg10, respectively, to afford the final tetradecasaccharide $Glc_3Man_9GlcNAc_2$ (Figure 1-3).[33] Unlike in the first phase of the pathway, the sugar donors on the lumenal side are the dolichylphosphate-linked monosaccharides Dol-P-Man and Dol-P-Glc.[35] These Dol-P-monosaccharides are synthesized by the enzymes Dpm1 and Alg5 on the cytoplasmic face of the ER membrane, and then flipped into the lumen by an unknown mechanism.[36-38] The underlying reasons behind why the cell switches to membrane-bound dolichyl-linked glycosyl donors in the ER lumen compared with the soluble nucleotide sugars in the cytoplasm are unclear, but presumably the scarcity of nucleotide sugar transporters in the ER membrane is a critical factor.

Although the structure of the core $Glc_3Man_9GlcNAc_2$ tetradecasaccharide is widely conserved across eukaryotes, several species of protists have been found to assemble only truncated forms of the glycan.[39,40] Recent studies have found that these truncations are due to a loss of certain sets of glycosyltransferases over the course of evolution.[41] For example, the causative agent of Chagas disease, *Trypanosoma cruzi*, produces only Dol-PP-$GlcNAc_2Man_9$ and is missing the glucosyltransferases Alg6, Alg8, and Alg10. In contrast, *Tetrahymena pyriformis* has lost all mannosylation function in the ER lumen, producing only Dol-PP-$GlcNAc_2Man_5Glc_3$. In an extreme case, *Plasmodium falicparum*, the cause of malaria infection in humans, is missing all the Alg glycosyltransferases except Alg7, thus generating only Dol-PP-GlcNAc and providing an explanation for the difficulty in identifying *N*-linked glycoproteins in this species.[41,42]

Glycan Transfer

After assembly of the oligosaccharide is complete, it is transferred to the side chain amide of acceptor proteins by the multimeric oligosaccharyl transferase (OT) complex. In *S. cerevisiae*, OT is composed of at least eight membrane-bound subunits (Ost3/6, Ost4, Stt3, Ost2, Wbp1, Swp1, Ost5, and Ost1) and genetic studies have determined that at least five of these (Ost1, Ost2, Stt3, Swp1, and Wbp1) are essential for survival (Figure 1-4).[43,44] Recent evidence indicates that multiple isoforms of the OT complex containing various combinations of subunits can exist within the same cell and may serve to modulate activity. For example, OT isoforms containing Ost1 and Swp1 are required for glycosylation of certain membrane proteins,[45] while the presence of either Stt3A or Stt3B impacts the kinetics and timing of glycosylation, promoting either co- or posttranslational catalysis.[46]

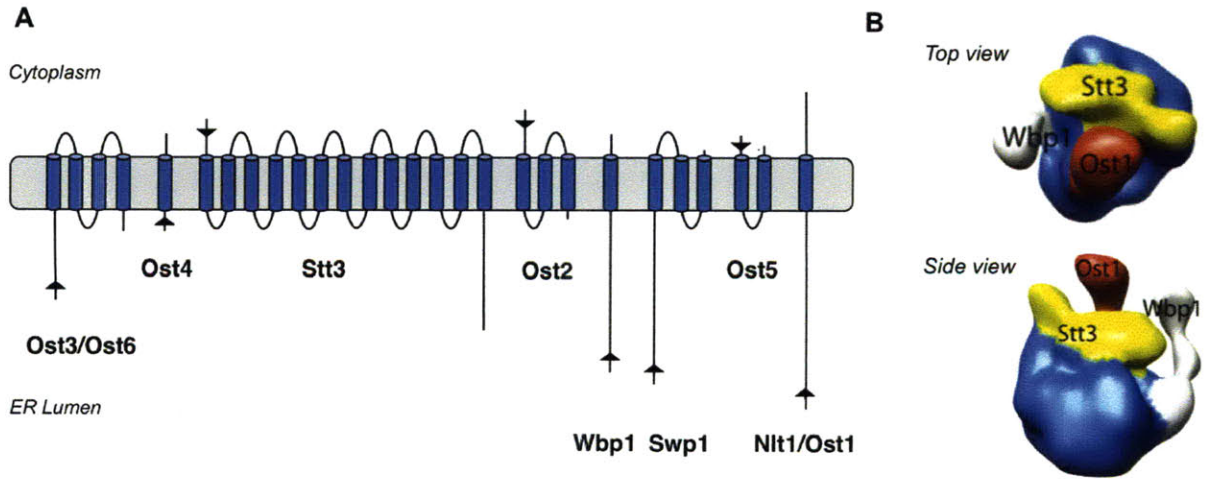


Figure 1-4: Subunits of the *S. cerevisiae* oligosaccharyl transferase complex. (A) Predicted topology of the individual subunits, where the arrows indicate the N-termini of the individual proteins; (B) Three-dimensional cryo-EM map of the OT complex, as reported by Li et al.[47]

Despite challenges associated with isolation and expression of the complex, several key studies have contributed toward understanding the specific roles of each subunit. The putative binding site of the dolichyl-linked oligosaccharide was assigned to Wbp1, as incubation of OT with Dol-PP-$GlcNAc_2$ was able to prevent alkyation of specific cysteine located within the Wbp1 subunit.[48] The Ost3/Ost6 proteins have been recently shown to exhibit protein-dependent oxidoreductase activity in vitro, suggesting a direct role for OT in protein folding.[49] In addition, a number of factors have emerged to suggest that Stt3 is the catalytic subunit of the OT complex. Stt3 is the most highly conserved subunit of the OT, as homologs of this protein are found in all eukaryotes and throughout the archaeal and bacterial kingdoms, and the Stt3 homolog in the bacterial pathogen *Campylobacter jejuni,* PglB, is sufficient to catalyze glycan transfer.[50] In addition, the eukaryotes *Leishmania major* and *Trypanosoma brucei* were found to contain only the Stt3 subunit of OT but are still capable of *N*-linked glycosylation.[51-53] Although structural studies of the entire OT complex are currently lacking, progress has been made to understand the proximity of each subunit to the others. To this end, the cryo-EM structure of the OT complex at 12 Å was recently reported (Figure 1-4B).[47]

The conformational requirements of the acceptor protein have been widely analyzed and do not appear to be very stringent.[44] However, not all asparagines found within the N-X-T sequon are glycosylated; in fact, it was recently estimated that roughly 35% of *N*-linked sequons that pass through OT are not modified.[44,54] Early work showed that OT does not glycosylate proteins that contain a proline in the X position of the N-X-S/T sequence,[55] and exhibits a modest preference for N-X-T sites compared with N-X-S.[56] In addition, sites within 12-14 residues of the N-terminus are not glycosylated, suggesting that the active site is 30-40 Å from the ER lumenal membrane.[57] The secondary structure of the glycosylation site has also been

found to be a determinant for catalysis, as the N-X-S/T sequon must be able to adopt an Asx-turn, which has been proposed to promote activation of the side chain amide nitrogen in the catalytic mechanism.[58,59]

Functional Roles of N-Linked Glycans

After glycosylation, the newly formed glycoproteins are processed by a series of glycosidases and glycosyltransferases in both the ER and Golgi to alter the core tetradecasaccharide, resulting the vast diversity of *N*-linked glycan structures observed in eukaryotes. These modifications include glycan trimming, as well as the addition of carbohydrates such as sialic acid, fucose, and *N*-acetylgalactosamine to form complex and branched structures.[1] The mature glycans are important for a number of biological functions, including promoting protein stability and rigidity, intracellular targeting, cell signaling, and the immune response. Mutations in almost every step of the glycan biosynthesis pathway lead to a group of debilitating and often fatal diseases that are collectively termed the congenital disorders of glycosylation (CDG), underscoring the importance of this protein modification to human health.[60]

Perhaps the most well-studied role for *N*-linked glycans in eukaryotes involves quality control for protein folding in the ER.[61,62] In the cotranslational pathway of *N*-linked glycosylation, newly synthesized proteins leaving the ribosome are translocated through the ER membrane by way of the Sec61 translocation channel and then presented to the OT complex. After glycan transfer, the nascent glycoproteins enter an elaborate quality control cycle that serves to ensure proper folding and utilizes all three branches of the core glycan (Figure 1-5).

The first step in this cycle involves the removal of two of the terminal glucose residues from the glycan core by glucosidases I and II.[63] The nascent glycoprotein then encounters two ER chaperones that recognize the monoglucosylated glycan, the membrane-bound calnexin and soluble calreticulin. These chaperones sequester the newly formed protein to prevent aggregation and misfolding, and also provide access to ERp57, an oxidoreductase that aids in disulfide bond formation. Once the protein is properly folded, glucosidase II removes the remaining glucose, which stimulates release of the protein from calnexin and calreticulin and progression through the remainder of the glycan-processing pathway in the ER and Golgi.

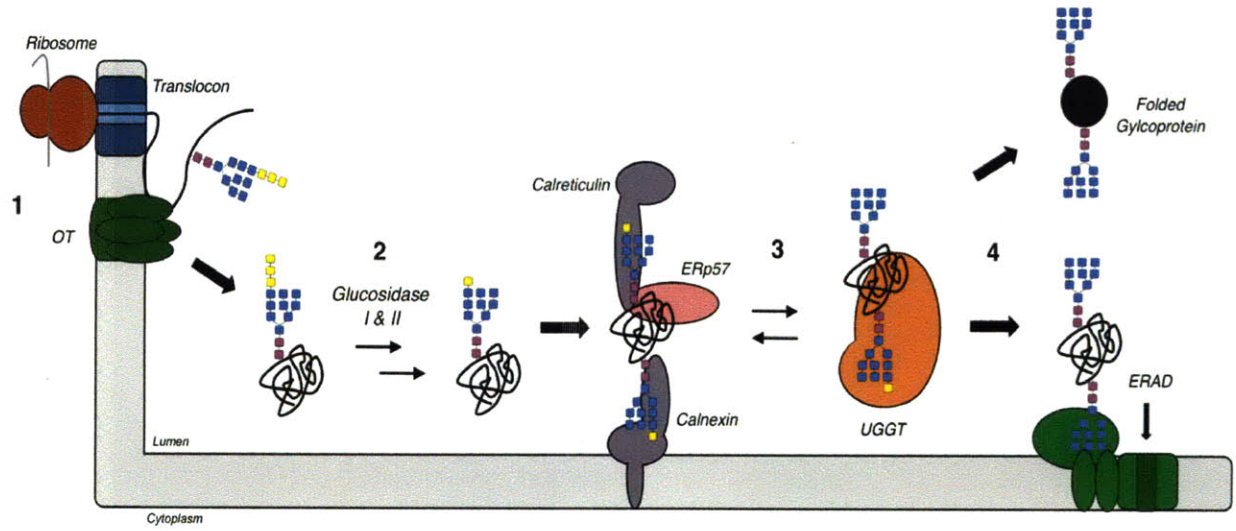


Figure 1-5: The calnexin/calreticulin pathway of quality control for protein folding. (1) Nascent proteins travel through the translocon and are glycosylated by OT; (2) The two terminal glucose residues are removed by glucosidase I and II; (3) The folding protein interacts with the chaperones calnexin and calreticulin and the ERp57 oxidoreductase. If misfolded, they are recognized by UGGT and cycled through the pathway again; (4) Defective proteins are demannosylated and ultimately sent for degradation through the ERAD pathway, while properly folded proteins exit the cycle.

However, if the protein is misfolded, it binds to the folding sensor UDP-glucose:glycoprotein glucosyltransferase (UGGT).[63,64] UGGT contains a hydrophobic N-

terminal domain that recognizes non-native protein structures, and a C-terminal glucosyltransferase domain that reglucosylates misfolded proteins, causing them to rebind calnexin and calreticulin. The iterative removal and addition of glucose by glucosidase II and UGGT persists until either the protein is folded or it is recognized by the α-1,2-mannosidases present in the ER, which remove terminal mannose residues from the glycan core.[64] Progressive mannose removal lowers the glycan affinity for calnexin/calreticulun and ultimately serves as a folding timer to interrupt the folding cycle and target the defective glycoprotein for degradation through the ER-associated degradation pathway (ERAD).[65]

In addition to their role in the quality control of protein folding, *N*-linked glycans perform a variety of other functions within eukaryotes. They serve as identity and localization tags to direct proteins to the proper cellular destination, and regulate mobility of proteins on or within the cell membrane through interactions with lattice-forming lectins, such as galectins.[66] *N*-linked glycans also impart structural rigidity to their targets and may provide protection from proteolysis, such as in the case of the heavily glycosylated lysosomal membrane proteins Lamp-1 and Lamp-2.[67,68] In addition, the surface *N*-linked glycans of cancer cells have been found to be highly branched compared with those of healthy tissues; exploitation of this difference may present a platform for the development of clinical biomarkers of disease.[69,70]

_N_-linked Glycosylation in Bacteria

Despite the enormous complexity and structural diversity of glycans that comprise the lipopolysaccharide (LPS) and capsular polysaccharide (CPS) of bacterial cell walls, it was believed for the past half century that bacteria were unable to glycosylate proteins.[71-73]

However, in the last decade, this dogma has been disproven with discovery of both *N*-linked and *O*-linked bacterial glycoproteins. In 1999, the first evidence of a general system of *N*-linked glycosylation was uncovered in the human mucosal pathogen *C. jejuni*.[74] The genes in this *pgl* locus (*p*rotein *gl*ycosylation) are homologous to specific genes in both the CPS or LPS pathways, yet mutations in this cluster resulted in a loss of immogenicity without affecting LPS or CPS biosynthesis, providing an initial link between *N*-linked glycosylation and host pathogenicity.[74,75] Structural characterization of the glycoproteins PEB3 and CgpA determined that the glycan was the heptasaccharide GlcGalNAc$_5$Bac, which forms a β-linkage to asparagine (Figure 1-6).[76,77] To date, over 65 glycoproteins of various biological functions have been characterized in *C. jejuni*, although new approaches to glycan analysis coupled with access to newly sequenced bacterial genomes suggest that number is sure to increase.[7]

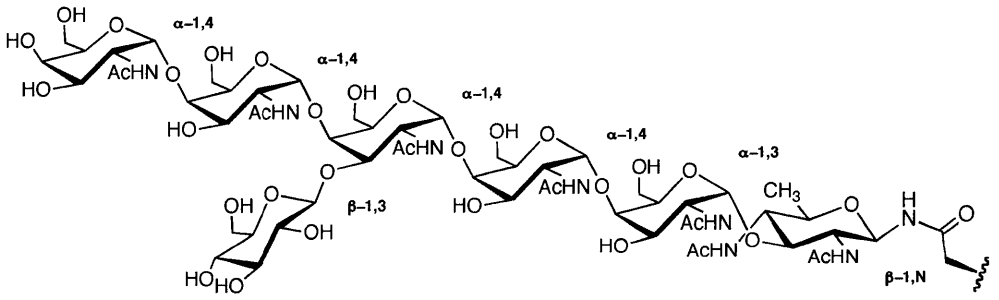


Figure 1-6: Structure of the *N*-linked GlcGalNAc$_5$Bac heptasaccharide identified in *C. jejuni*.

Glycan Assembly

The *N*-linked glycosylation pathway in *C. jejuni*, shown in Figure 1-7, bears resemblance to the first half of the dolichol pathway in *S. cerevisiae* (Figure 1-2). The glycan is assembled on the cytoplasmic face of the membrane, and involves stepwise elaboration onto an

polyisoprenyl-diphosphate carrier by a series of glycosyltransferases that rely on nucleotide sugar donors. However, although the overall architecture of the pathway is similar, several differences distinguish *N*-linked glycosylation in eukaryotes and bacteria. For one, the isoprenyl carrier in bacteria is undecaprenol, a long chain unsaturated polyisoprene composed of 11 units that is the most abundant isoprene in bacterial membranes.[78] In addition, the first sugar in the *C. jejuni N*-linked glycan is *N,N'*-diacetamidobacillosamine (Bac), rather than the GlcNAc residue found in eukaryotic glycans. In general, prokaryotes contain a variety of unique, highly-modified sugars not commonly observed in eukaryotes.[79,80] Originally discovered in *Bacillus subtilis*, Bac is found in a number of bacterial glycans including the LPS of *Vibrio cholerae*, the CPS of *Acinetobacter lwofii*, and the pilin of *Neisseria meningitidis*.[81-84]

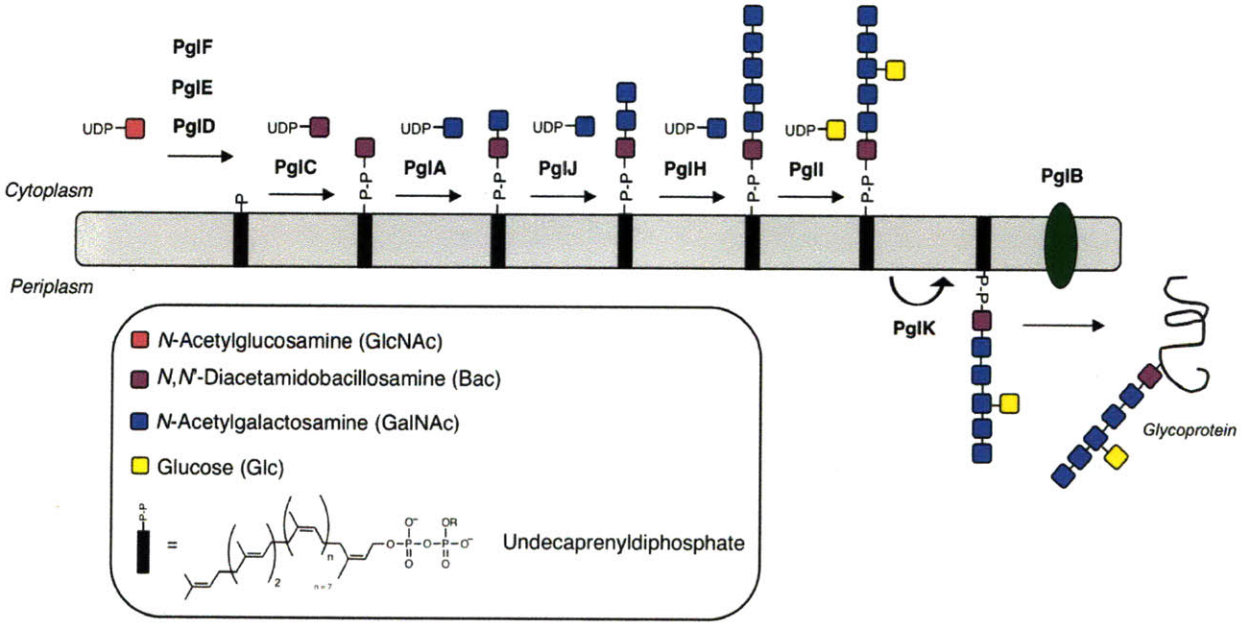


Figure 1-7: *N*-linked glycosylation pathway in *C. jejuni.*

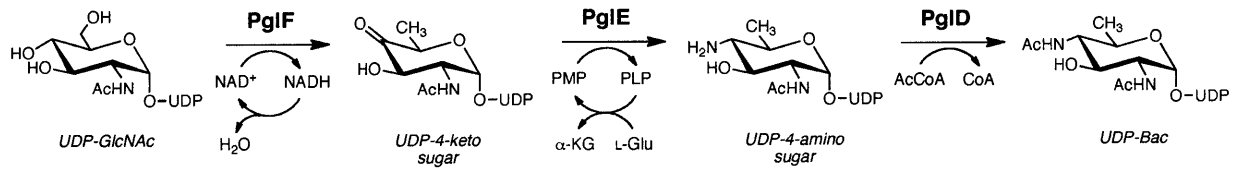


Figure 1-8: Biosynthesis of UDP-Bac from UDP-GlcNAc by the *C. jejuni* enzymes PglF, PglE and PglD.

The *N*-linked glycosylation pathway in *C. jejuni* is initiated by the biosynthesis of the UDP-Bac by the enzymes PglF, PglE and PglD (Figure 1-8). Starting from UDP-GlcNAc, the dehydratase PglF generates the UDP-4-keto intermediate via an NAD^+-promoted hydride transfer and facilitated elimination, followed by formation of the UDP-4-amino sugar by the PLP-dependent aminotransferase PglE.[85] PglD then catalyzes the acetylation of UDP-4-amino using AcCoA to afford UDP-Bac.[86] The phosphoglycosyltransferase PglC initiates carbohydrate transfer to the undecaprenylphosphate carrier by installing Bac-1-P, succeeded by the addition of the first and second α-1,3 and α-1,4-linked GalNAc residues by PglA and PglJ, respectively.[87-89] Analysis of the polyisoprene specificity of these enzymes revealed that the α-unsaturation and the *cis*-double bond geometry of undecaprenol was more important than small changes in the polyisoprene length, highlighting the significance of the specific polyisoprene structure to the enzymes in the glycosylation pathway.[90] The trisaccharide is then elaborated through the action of the polymerase PglH, which installs the next three α-1,4-linked GalNAc residues.[91] Detailed kinetic examination revealed that PglH uses a single active site to carry out three galactosyltransfer reactions, and exhibits an increasingly stronger affinity for its product as the glycan size lengthens, which serves to halt catalysis after the third transfer. The assembly of the heptasaccharide is completed with the addition of a branching β-1,3 linked glucose residue

by the membrane protein PglI. The entire *pgl* locus along with two *C. jejuni* perisplasmic proteins, AcrA and PEB3, have been inserted into *Escherichia coli* and the system was shown capable of glycosylation, indicating that all of the necessary components of the pathway were present.[92] Interestingly, these enzymes can all be combined with the obligate cofactors and cosubstrates to generate the full length Und-PP-BacGalNAc$_5$Glc in vitro, which represents the first complete synthesis of a polyisoprenyl-linked oligosaccharide.[86,89]

In addition to *C. jejuni*, evidence of *N*-linked glycosylation pathways in other species of bacteria is slowly beginning to emerge. *Wolinella succinogenes* was shown to contain a *pgl* gene cluster, although the glycan structure as well as the identity of possible target proteins is still unclear.[93] Structural analysis of the HmcA protein from the sulfate-reducing bacterium *Desulfovibrio gigas* revealed the presence of a trisaccharide linked to asparagine, but the exact composition of this glycan remains unknown.[94] Several species of *Helicobacter*, including *H. canadensis* and *H. pullorum*, also possess *pgl* gene clusters. Recent biochemical studies have shown that *H. pullorum* contains two isoforms of the *pglB* OTase gene; however, only one of these genes is responsible for transfer of a linear *N*-linked glycan comprising five hexosamine sugars, while the role of the second isoform is still under investigation.[95] Bioinformatic analyses of newly sequenced bacterial genomes suggests the presence of *pgl* genes in over 49 bacterial species, although biochemical characterization of these pathways is scant.[7]

Interestingly, *O*-linked pilin glycosylation in both *Neisseria meningitides* and *Neisseria gonorrhoeae* is believed to strongly resemble *N*-linked glycosylation in *C. jejuni* (Figure 1-9).[96,97] In the *N. gonorrhoeae* pgl (*p*ilin *gl*ycosylation) pathway, genetic studies suggest that the enzymes PglB, PglA and PglE are responsible for assembly of a trisaccharide onto an undecaprenyldiphosphate-linked carrier, which is then flipped into the periplasm and transferred

to the hydroxyl side chain serine atom by the oligosaccharyl transferase PglO. This pathway differs from most *O*-linked glycosylation pathways in eukaryotes and bacteria in that the glycan is transferred to the serine residue in an *en bloc* manner, rather than the individual transfer of each sugar directly onto the protein from a nucleotide diphosphate donor. The glycan structure is believed to be Gal_2Bac, though the stereochemistry of the Bac moiety has not yet been assigned.[83] The PglO OTase does not appear to require a specific glycosylation sequence in the acceptor protein, and also does not bear close sequence homology to the *C. jejuni* OTase PglB.[97,98] Detailed biochemical characterization of this pathway is still required to understand the basis of glycosylation specificity.

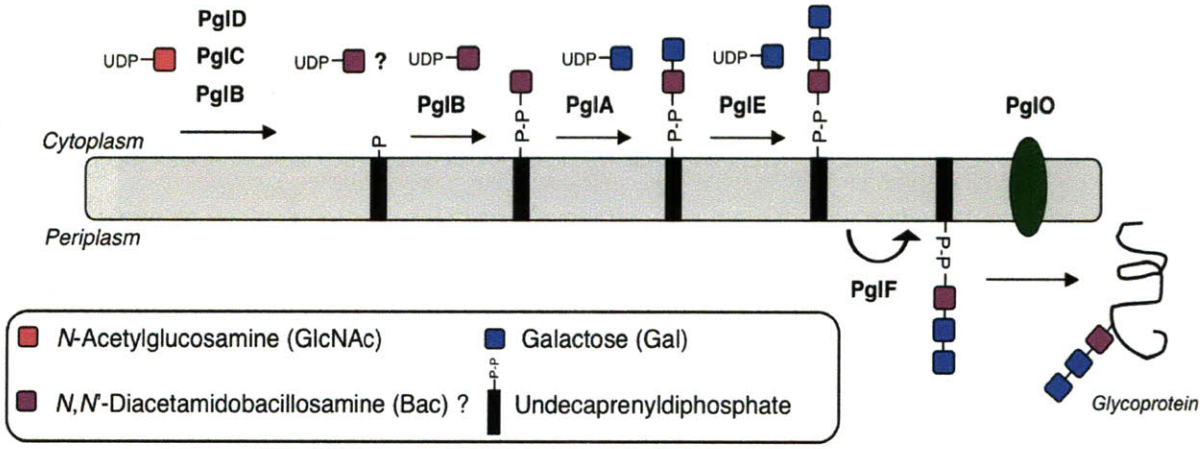


Figure 1-9: Proposed pathway for *O*-linked pilin glycosylation in *N. gonorrhoeae*.

Glycan Transfer

In contrast to the multimeric OT complex found in most eukaryotes, the transfer of the bacterial *N*-linked glycan is carried out by a single protein, the Stt3 homolog PglB.[50,92] PglB is predicted to consist of a large N-terminal domain made up of 9-13 transmembrane helices, and a periplasmic C-terminal domain that contains the highly conserved WWDXGX sequence (Figure

1-10). A preliminary crystal structure of the C-terminal domain of PglB was recently reported, and although this domain is not sufficient to catalyze oligosaccharide transfer, analysis of this structure coupled with evaluation of genetic point mutations and comparison with other PglB homologs suggest the importance of the two additional protein motifs, MXXI (residues 568-571) and XXD (residues 52-54), for catalysis.[99]

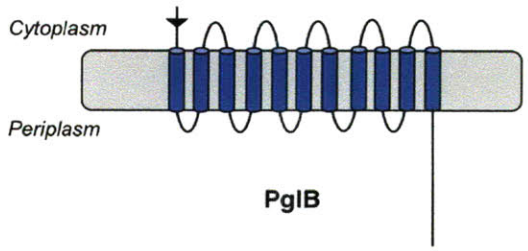


Figure 1-10: Predicted membrane topology of the *C. jejuni* OTase, PglB. The size of the loops between the transmembrane domains is variable and has not yet been determined.

Recent studies have shown that PglB is able to glycosylate fully folded proteins and requires the extended glycosylation sequence D/E-X_1-N-X_2-S/T, in which an acidic residue occupies the -2 position.[100,101] Experiments to define the optimal glycosylation sequence involved testing a library of peptides in which the residues at both X positions were varied; PglB was found to glycosylate peptides containing a bulky residue at the X_1 position and a smaller hydrophobic residue in the X_2 position with the greatest efficiency.[102] The extremely high affinity that PglB exhibits for these short acceptor peptides (K_m = 1 μM for DQNAT) suggest that PglB does not require a specific tertiary structure for binding its fully folded protein substrate.[102] In addition, structural analysis of glycosylated proteins PEB3 and AcrA have confirmed that the glycosylation sequence must be in a fairly exposed region of the protein in

order to facilitate access by PglB.[103,104] PglB has also been shown to be more promiscuous than the eukaryotic OT in regard to the glycan structures that it is capable of transferring to protein. Experiments in which PglB was substituted for the bacterial LPS O-antigen ligase demonstrated that PglB was able to transfer distinct O-antigens from both *E. coli* and *P. aeruginosa* onto asparagine residues, highlighting its increased substrate flexibility.[105] However, the relative efficiency of glycosylation for these alternate substrates has not yet been biochemically determined.

Biological Significance of Bacterial N-Linked Glycans

To date, over 60 distinct glycoproteins have currently been identified in *C. jejuni*.[7] These proteins are associated with a wide variety of biological pathways within the cell, suggesting that *N*-linked glycans may play a role in a number of important cellular functions. Initial genetic studies in which the *N*-linked glycosylation pathway was disrupted have resulted in impairments in cell adhesion and host colonization, establishing a strong tie between *N*-linked glycans and virulence. For example, early studies of *C. jejuni* glycoproteins indicated that they were highly immunogenic when cross-reacted with animal antisera.[74] In addition, genetic mutants of *C. jejuni* that are impaired in glycosylation showed a reduced ability to adhere to and invade human intestinal epithelial cells, as well as decreased colonization of the intestinal tracts of mice and chicken.[75,106] Although details about the interactions between *C. jejuni* and the mammalian immune system are largely unclear, a recent study revealed that the *C. jejuni N*-linked glycan is recognized by the human macrophage galactose-type lectin, a dendritic cell receptor that plays an important role in intracellular signaling.[107] When glycosylation was impaired, *C. jejuni* was

found to stimulate formation of the cytokine IL-6, suggesting that glycosylation of specific proteins might help the organism evade the host immune response.

Other roles for the *N*-linked glycans of *C. jejuni* are slowly beginning to emerge. The VirB10 protein, a component of the type IV secretion system that contains two glycosylation sites, was found to be incapable of properly forming a complex with other secretion system proteins when glycosylation of a key asparagine was prevented, implying that the glycan may be important in modulating protein-protein interactions.[108] Analysis of the recent NMR structure of the glycoprotein AcrA, part of the *C. jejuni* multidrug efflux pump, shows that the *N*-linked glycan adopts a rigid, rod-like structure that appears to fold back over part of the protein (Figure 1-11).[104] The orientation and rigidity of the glycan indicate that it may serve to enhance protein stability and provide protection from proteolysis. *N*-linked glycans have also been associated with protection of *C. jejuni* from osmolytic stress. A recent study revealed that *C. jejuni* is able to produce free oligosaccharide in response to variations in osmolarity of the external environment.[109] While the mechanism for oligosaccharide release is still unclear, early studies implicate proteins in the Pgl pathway, particularly PglB, in the accumulation of these structures in the *C. jejuni* periplasm.

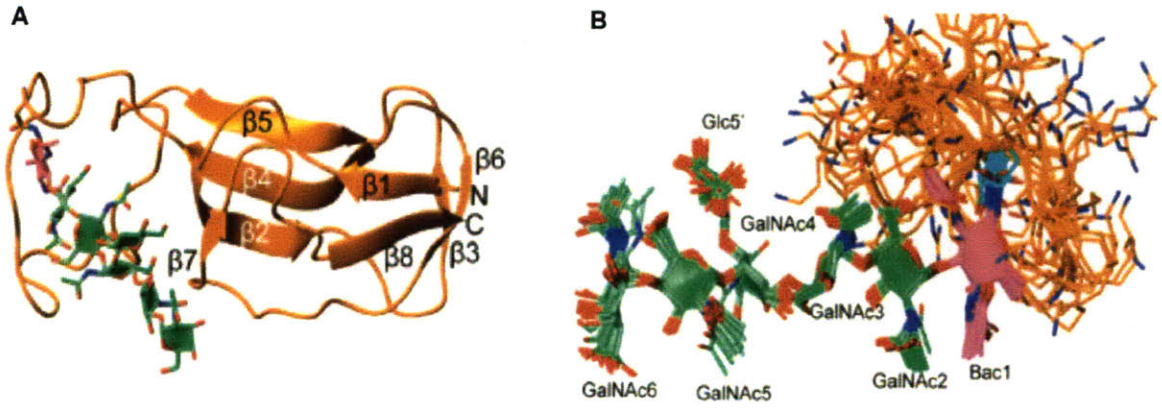


Figure 1-11: NMR structure of the glycoprotein AcrA, modified with the $GlcGalNAc_5Bac$. Bac is colored pink, while the other residues are shown in green. (A) Representative structure of the glycoprotein; (B) Structural ensemble of the glycoprotein. Figures adopted from Slynko et al.[104]

N-linked Glycosylation in Archaea

Structural Diversity of N-Linked Glycans

First described by Carl Woese in 1977, archaea are single-celled organisms that share many common features with both bacteria and eukaryotes, but also possess several characteristics unique to their kingdom.[110] Like bacteria, archaea are prokaryotes that lack internal organelles and contain a single, circular genome. However, the manner in which genetic information is organized, replicated, and transmitted over the course of the archaeal life cycle is strikingly similar to eukaryotes. Archaea are distinct from both bacteria and eukaryotes in several respects; for example, archaeal membranes are composed of ether-linked lipids, in contrast to the ester-linked phospholipids found in bacterial and eukaryotic membranes.[111] In addition, archaea inhabit a wide variety of environments compared with other life forms, from extreme habitats such as hot springs, deep sea thermal vents, and highly saline lakes, to more familiar regions such as soil, seawater, and the mammalian digestive tract.[112]

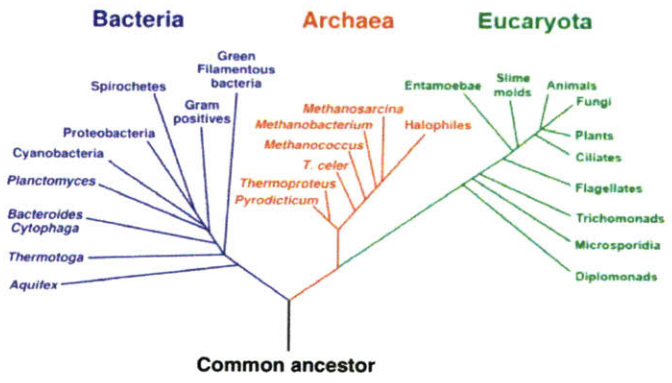


Figure 1-12: Phylogenetic tree depicting the three kingdoms of life.

The first archaeal glycoproteins were discovered in 1976 in the extreme halophile *Halobacterium salinarium*.[113] In an intriguing study, the structural analysis of the closely related species *Halobacterium halobium* revealed that the S-layer protein of the organism was modified with two very distinct types of *N*-linked glycans. The first was determined to be a pentasaccharide made up of sulfated glucose and glucuronic acid (GcA) residues (Figure 1-13A).[114,115] The second glycan discovered is much larger and more complex, with similarities to the glycosaminoglycans that are found on connective tissues in higher eukaryotes. It involves the β-linkage of a sulfated GalNAc to the asparagine of the protein, followed by between 10-15 repeating units of a branched pentasaccharide containing sulfated glucose, galacturonic acid (GalA) and galactosaminuronic acid (GalNAcA) residues, as well as a Gal in the furanose configuration (Figure 1-13B).[116,117] These long glycans were found to coat the outer surface of *H. halobium*, forming an acidic, two-dimensional lattice that surrounds the organism.[115]

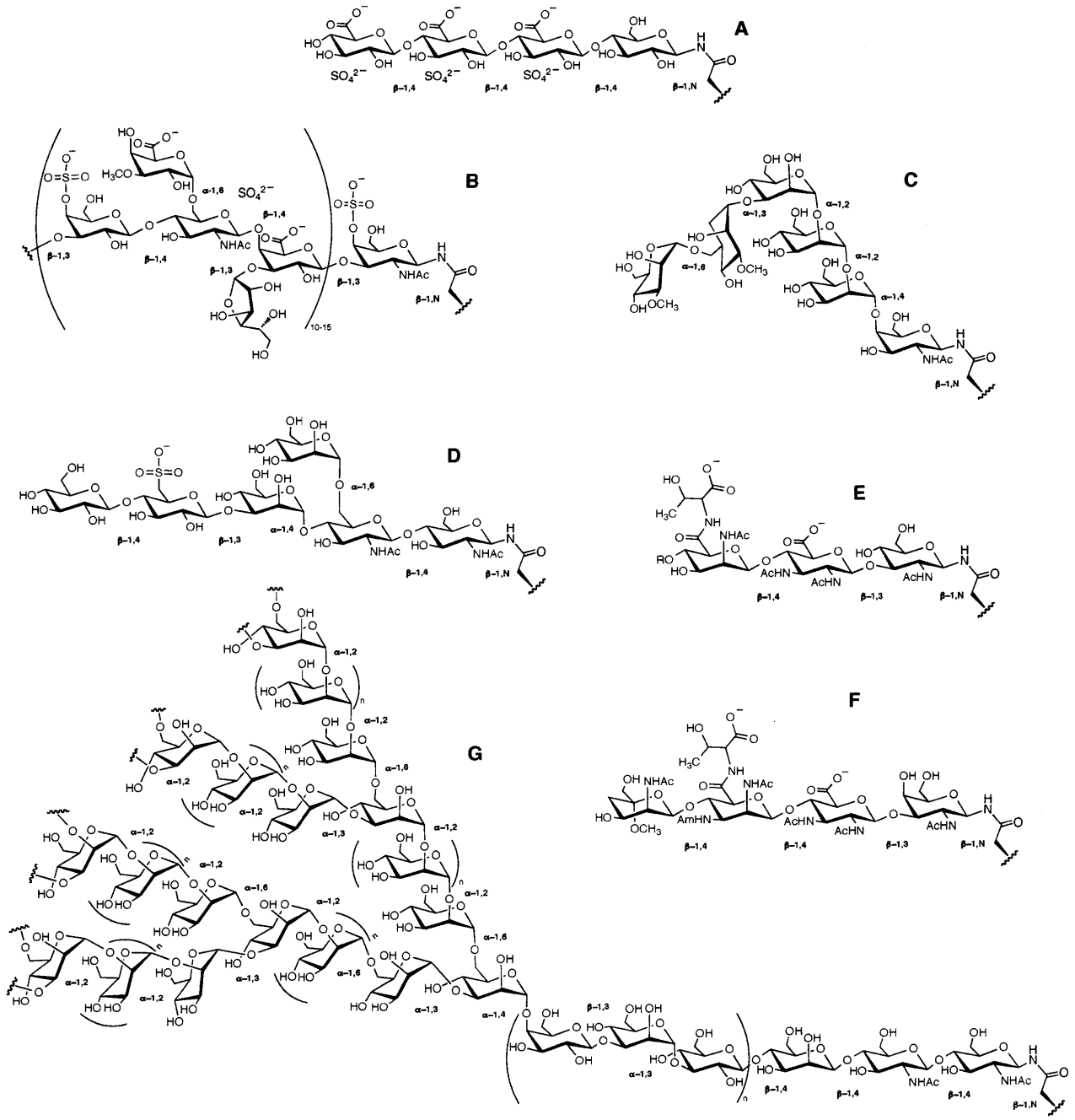


Figure 1-13: Structures of representative archaeal *N*-linked glycans. (A) short glycan, *H. halobium*; (B) longer glycan, *H. halobium*; (C) *M. fervidus*, the two residues at the nonreducing end are either Man or Glc; (D) *S. acidocaldarius*; (E) *M. voltae*, the R indicates the presence of an unidentified hexose in some strains; (F) *M. maripaludis*; (G) *T. acidophilum*, the order of the residues near the reducing end are unclear. Unusual sugar moieties are colored in blue; see text for more detail and specific references.

Since this discovery, many other *N*-linked glycoproteins have been identified across the archaeal kingdom. These glycans display an extraordinary structural variety, which may stem from the diverse habitats of the archaeal species that produce them. The S-layer protein of the hyperthermophilic methanogen *Methanothermus fervidus* is modified with an *N*-linked hexasaccharide, in which a GalNAc is attached to the Asn residue of the protein (Figure 1-13C).[118] The glycan contains three mannose residues followed by two methylated glucose or mannose residues at the nonreducing end. Another methanogen, the mesophile *Methanococcus voltae*, was found to contain glycosylated S-layer proteins and flagella, all labeled with the trisaccharide GlcNAc-β-1,3-GlcNAc(3NAc)A-β-1,4-ManNAc(6Thr)A (Figure 1-13E).[119] The terminal ManNAcA residue of this glycan is linked to a threonine through an amide bond at the C6" position. However, in some strains of *M. voltae*, the trisaccharide is further elaborated with an unknown hexose in the fourth position.[120] In *Methanococcus maripaludis*, the flagella are modified with a linear tetrasaccharide that is β-linked to the asparagine through a GalNAc residue (Figure 1-13F).[121] GlcNAc(3NAc)A is ligated to the proximal GalNAc sugar, followed by a highly modified mannosaminuronic (ManNAcA) residue that contains an acetamidino group at the C3" position and a threonine at the C6" carbon, similar to the *M. voltae* glycan. In addition, the terminal sugar of this oligosaccharide is the unique 2-acetamido-2,4-dideoxy-5-*O*-methyl-hexos-ulo-1,5-pyranose, which represents the first reported example of an aldulose in an *N*-linked glycan structure.[121]

In addition to the characterization of the methanogen-derived glycans described above, two other archaeal *N*-linked glycans have been annotated in detail. The cytochrome $b_{558/566}$ protein from the thermophile *Sulfolobus acidocaldarius*, which grows at optimal conditions of 75-80 °C and pH 2-3, is modified with a hexasaccharide linked through a GlcNAc that contains

a 6-sulfoquinovose residue (Figure 1-13D).[122] *Thermoplasma acidophiluim* produces a complex, highly branched *N*-linked glycan composed mainly of mannose that is attached to a surface membrane protein through a GlcNAc-Asn linkage (Figure 1-13G).[123] Copies of this large glycan encapsulate the cell membrane and are proposed to shield the organism from its harsh surroundings. Initial studies of other archaeal glycoproteins have been reported, including the those from the hyperthermophile *Pyrococcus furiosus* and the halophile *Haloferax volcanii*, although detailed structures of the respective glycans have not yet been elucidated.[8,124] For more information about proposed archaeal *N*-linked glycans, the reader is referred to a review by Eichler.[125]

Glycan Assembly

Unlike the analogous pathways in eukaryotes and bacteria, many of the details of *N*-linked glycan assembly in archaea are still unknown. A recent analysis of the nearly 60 sequenced archaeal genomes indicates that all but two (*Aeropyrum pernix* and *Methanopyrus kandleri*) contain an *stt3* homolog, which suggests that in contrast to bacteria, *N*-linked glycosylation in archaea may be widespread.[126] However, difficulties with genetic manipulation and large-scale growth of these organisms have hampered genetic and biochemical studies. Despite these challenges, progress toward uncovering the biosynthesis of archaeal *N*-linked glycans has begun to accelerate, particularly in the last five years.

Recent genetic studies in the methanogens *M. voltae* and *M. maripaludis* and the halophile *H. volcanii* have provided the first glimpses of the glycan assembly pathways in archaea; the current working models of *N*-linked pathways in these organisms are summarized

in Figure 1-14. In all three models, glycan assembly is believed to take place on the cytoplasmic face of the cell membrane and is catalyzed by a series of Agl glycosyltransferases (*a*rchaeal *gl*ycosylation), after which the glycan is flipped to the exterior of the cell and transferred to asparagine residues of acceptor proteins by the Stt3 homolog AglB.[127] Early studies in *H. halobium*, *H. volcanii,* and *T. acidophilum* determined that archaeal membranes contain both the monophosphate and diphosphate forms of dolichol, where the dolichol is only 11-12 isoprene units in size.[128-130] The presence of both forms of dolichylphosphate is reminiscent of the eukaryotic glycosylation pathway, in which Dol-P-sugars are utilized as the glycosyl donors for at least some of the steps of glycan assembly. It is unclear whether archaea also rely on Dol-P sugars for assembly of the *N*-linked glycan. The partial characterization of a putative dolichyl-phosphomannose synthease (DPM) from *T. acidophilum* lends support to this hypothesis;[131] however, further biochemical evidence is still required to understand the individual roles of the each dolichylphosphate compound.

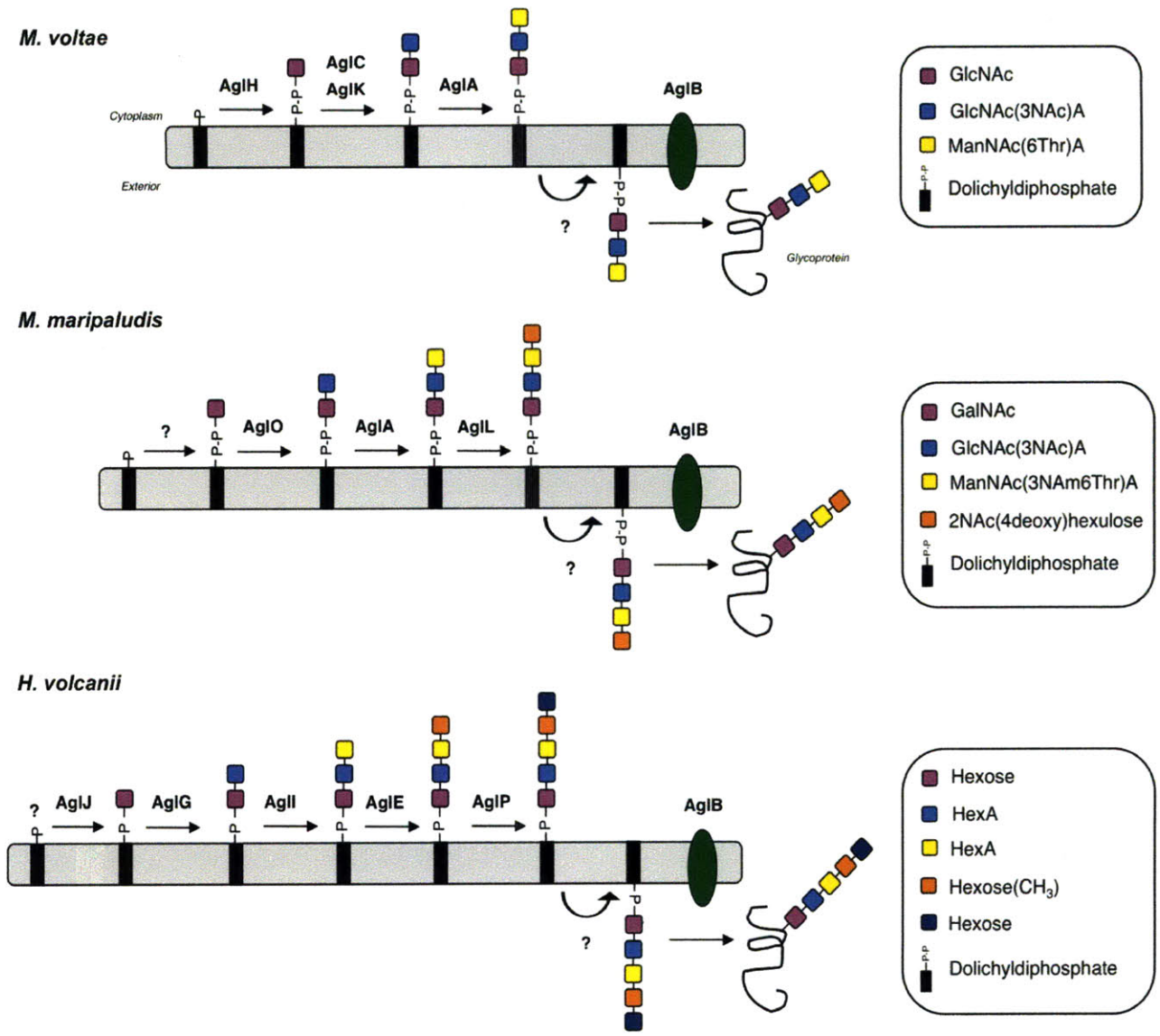


Figure 1-14: Current models of *N*-linked glycosylation in *M. voltae*, *M. maripaludis*, and *H. volcanii* based on recent genetic studies. The exact structure of the *H. volcanii N*-linked glycan has not yet been determined, but data suggest that it is a pentasaccharide as indicated.

As shown in Figure 1-14, the first step of the *N*-linked glycosylation pathway in these selected archaea involves the transfer of a sugar 1-phosphate onto the dolichyl-linked donor by the putative phosphoglycosyltransferase AglH (AglJ in *H. volcanii*). Strong evidence for this

initial step was provided in a recent study in which the *M. voltae algH* gene was able to complement a conditionally lethal mutant of the analogous phosphoglycosyltransferase gene in *S. cerevisiae*, *alg7*.[132] This finding suggests that in *M. voltae*, the glycosyl donor for AglH is UDP-GlcNAc and the polyisoprenyl carrier is dolichyldiphosphate. In the related methanogen *M. maripaludis*, a similar enzyme is believed to transfer GalNAc 1-P to Dol-P, though the candidate gene has not yet been identified. An alternate model for formation of the Dol-PP-GalNAc intermediate in *M. maripaludis* could involve initial formation of Dol-PP-GlcNAc by a still undiscovered AglH homolog, followed by the epimerization of Dol-PP-GlcNAc to Dol-PP-GalNAc. However, this hypothesis has yet to be supported with genetic or biochemical evidence, although an epimerase that catalyzes the conversion of Und-PP-GlcNAc to Und-PP-GalNAc in *E. coli* was recently described along with the proposal of a homologous candidate epimerase in *M. maripaludis*.[133,134]

The remainder of the steps involved in glycan assembly remain unclear. Initial genetic deletion studies suggest two different paths for the transfer of the second sugar GlcNAc(3NAc)A in both *M. voltae* and *M. maripaludis*, though this finding has not been verified biochemically.[120,135] Transfer of the third sugar is believed to be carried out by AglA or AglI; interestingly, in both the *Methanococcus* species, this sugar is a mannuronic acid derivative modified with a threonine residue at the C6" carboxylate. After completion of the glycan, it is flipped across the cell membrane into the external environment, where an OTase presumably catalyzes its transfer to acceptor proteins. Evidence for the localization of oligosaccharide transfer to the external face of the cell membrane comes from the treatment of *H. salinarium* with bacitracin, which inhibited formation of sulfated glycoproteins.[136] Bacitracin, a small molecule that is unable to traverse the archaeal membrane, binds to the

phosphate groups of free Dol-PP that is liberated by the OTase after glycosylation, thus preventing recycling of the dolichol.[137] However, despite this evidence, a candidate flippase gene has been yet to be identified in archaea. Further work involving detailed genetic and biochemical studies is required to answer the many questions that remain about the biosynthesis of these intriguing glycans.

Glycan Transfer

The archaeal Stt3 homolog AglB is predicted to have a similar overall topology to the bacterial PglB, containing both a large N-terminal region composed of 11-13 transmembrane domains and an extracellular C-terminal domain that bears the signature OTase WWDXGX sequence (Figure 1-10). Recently, the X-ray crystal structure of the C-terminal domain of AglB from *P. furiosus* was reported, providing the first insight into the structural organization of a segment of an Stt3 homolog.[124] Although it is unclear if this structure represents a biologically relevant form of the enzyme, as the C-terminal domain alone was unable to catalyze oligosaccharide transfer, this study revealed that the WWDXGX motif is located in a central β-helix in close proximity to a highly conserved DXXK motif. Initial biochemical characterization of the full length AglB using crude mixtures of membrane-associated lipids suggests that the enzyme alone is sufficient to carry out glycosylation, and that mutations in either the WWDXGX or DXXK motif impair glycosylation activity.[124] These results are consistent with genetic studies in *M. voltae* and *M. maripaludis*, in which mutations in *aglB* gene resulted in flagella with decreased glycosylation and impaired motility.[127,138]

In contrast to the bacterial system, the archaeal OTase AglB does not require an extended glycosylation sequence beyond the N-X-S/T sequon defined for eukaryotic glycoproteins. However, a comparative analysis of the amino acid sequence around the glycosylation site of confirmed archaeal glycoproteins indicates a slight preference for Ala and Gly at the X_{+1} position and Ser or Thr at the X position, where X_{-1}-N-X-S/T-X_{+1}.[139] As AglB is predicted to glycosylate proteins in a posttranslationally, it is proposed that the target glycosylation site should be located in a flexible, exposed region of the acceptor protein as described for the protein substrates of PglB; however, structural analysis of glycoproteins is required to confirm this hypothesis.

Functional Role of N-linked Glycans

As archaea populate a wide range of habitats from the extreme to the more familiar, it is not surprising that the structures of their respective *N*-linked glycans mimic this diversity. Archaea already possess a variety of biological adaptations to equip them for life in chemically harsh environments. For example, archaea exhibit a unique composition of membrane lipids, such as diethers, tetraethers, and macrocyclic diethers, which results in the decreased ion permeability and planar dynamic capacity of their membranes.[140] In addition, thermophilic and halophilic archaea employ a number of strategies to enhance protein stability at high temperature and salt concentrations, including increased ion pairing and hydrophobic interactions, burying polar contacts into the interior of proteins, and decreased entropy of unfolding.[141-143]

The varied structures of *N*-linked glycans identified in archaea to date likely serve to equip these species for optimal growth and survival in their diverse surroundings. For example, the large, highly branched glycan on the surface of the hyperthermophile *T. acidophilum* is predicted to protect the organism from the heat and high acidity of its environment by forming an extensive hydrogen-bonded shell around the organism, providing external strength and rigidity and limiting access of certain ions and water molecules to the cell membrane.[123] In bacteria, these functions are provided by structures such as the LPS and CPS, but as archaea do not have cell walls, it is hypothesized that perhaps the surface *N*-linked glycans in species such as *T. acidophilum* and *H. halobium* impart similar protective features. Another important role for archaeal N-linked glycans is related to cell motility. In the methanogens *M. voltae* and *M. maripaludis*, the protein subunits that comprise the flagella are heavily modified with negatively charged *N*-linked glycans (Figures 1-13E and 13F). Genetic mutations that disrupt the glycosylation pathway in these organisms have resulted in impaired motility; in fact, in *M. maripalidus* cells, a direct relationship between glycan size and swimming capability has been established.[134] It is unclear exactly how glycans contributes to the proper function of flagella, although the extensive glycosylation may sheath the structure with negative charge and possibly impart stability to interactions amongst the individual protein subunits (Figure 1-15). Further studies are required to gain a deeper understanding of the roles of *N*-linked glycans in archaea.

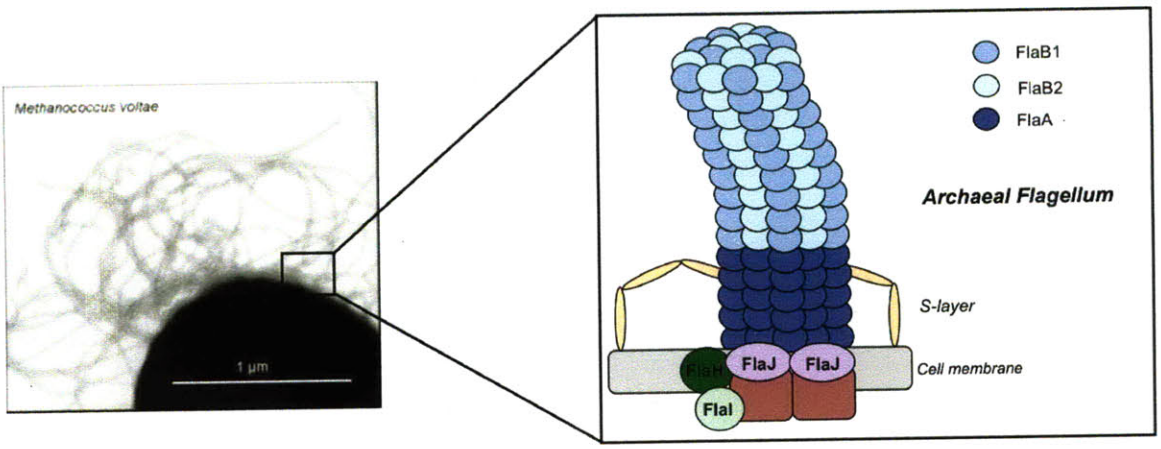


Figure 1-15: Composition of the *M. voltae* flagella. Each of the Fla proteins depicted in blue is modified with between 1-4 glycans. Figures adapted from Jarrell et al.[127,144]

Role of *N*-linked Glycosylation in Viruses

In recent years, the importance of *N*-linked glycosylation in virus biology has begun to emerge, revealing the many ways in which glycans contribute to viral survival and infection. Human viruses rely on the host cell glycosylation pathway in nearly every step of the viral life cycle, from host cell recognition, viral replication, protein trafficking and virion packaging.[1] Viruses hijack the host cell glycosylation pathways to take advantage of the many benefits experienced by eukaryotic proteins, including access to the host cell protein folding and quality control machinery. Nearly all of the viruses studied thus far have utilized the calnexin/calreticulin cycle to aid in protein folding and trafficking.[145] In addition, the rapid evolution of viral particles often results in an increased number of glycosylation sequences, introducing diversity and complexity into the viral glycan profile that can ultimately complicate treatment of infection.

The influenza virus is a common human pathogen of the upper respiratory tract and a member of the orthomyxovirus RNA virus family; in addition, it is perhaps the most well

studied virus with respect to glycosylation. The virus contains two important capsular coat glycoproteins, hemagglutinin (HA) and neuraminidase (NA), that utilize *N*-linked glycans for a number of important functions such as receptor binding and infection. The dominant surface protein is HA, which is modified with a heterogeneous, complex glycan that depends on the viral strain but is indicative of extensive processing in the Golgi.[146,147] HA is heavily glycosylated, containing between 5 and 11 glycosylation sites depending on the strain, all of which are clustered at the globular head of the protein and are important for interaction with a key sialic acid residue on the host cell receptor. Upon viral replication and release, NA acts to cleave this sialic acid moiety to promote diffusion of the virus. If blocked, the newly formed viral particles rebind the host cell receptor and form large aggregates. Increased glycosylation of HA to improve viral binding requires enhanced activity of NA to mediate viral release, thus demonstrating the stategic interplay of these two glycoproteins for effective influenza infection.

Another important example of viral glycosylation is the human immunodeficiency virus (HIV), a highly mutagenic RNA virus from the retrovirus family. The virus is surrounded by the multiple copies of the envelope protein gp120, one of the most heavily glycosylated structures found in nature.[148] The structure of the *N*-linked glycan coating gp120 was only just recently characterized and found to be a high oligomannose structure, suggesting that this glycan is only minimally processed after transfer to gp120 by OT.[148] A recent analysis of global HIV strains indicated that the gp120 protein contains between 18-33 possible *N*-linked glycosylation sites, with average of 25 glycans per protein.[149] The vast density of carbohydrates forms a large, encompassing structure called the glycan shield, which is postulated to protect the virus from recognition and degradation from the host immune system. In particular, the glycan shield provides protection from the 2G12 neutralizing antibody, which has been shown to recognize

the virus particle and render it inactive.[150] Rapid viral evolution introduces multiple amino acid substitutions into the gp120 sequence, changing the number and organization of the glycan sites and thus allowing the virus to persist in spite of increasing pressure from host antibodies.[145]

N-linked glycosylation has also been found to play an important role in the infection and stability of other viruses as well, including the West Nile, Ebola, SARS, and Hantaan viruses.[145] However, further work is required to uncover the mechanisms of host recognition and viral fusion. In general, a deeper understanding of these viral glycoproteins may provide new opportunities for development of antiviral therapeutics.

Conclusions

N-linked glycosylation plays an important role in many biological processes across eukaryotic, bacterial, and archaeal kingdoms. Although the overall architecture of each pathway shares many features, a comparison of the details of glycan assembly and transfer in each kingdom reveals a number of intriguing differences. Table 1-1 summarizes the defining characteristics of each of these pathways.

Table 1-1: Comparison of *N*-linked glycosylation across the three kingdoms

Characteristic	Eukaryotes	Bacteria	Archaea
Site of glycan assembly	ER membrane	Inner Cell Membrane	Cell Membrane
Isoprene carrier	Dol-PP (14-21 isoprene units)	Und-PP (11 isoprene units)	Dol-P and Dol-PP (11 isoprene units)
Glycosyl donor	NDP-sugars, Dol-P-sugars	NDP-sugars	NDP-sugars, Dol-P-sugars (?)
Asn-linked sugar	GlcNAc	Bac	GlcNAc, GalNAc, Glc
Glycan diversity across kingdom	Initially conserved ($Glc_3Man_9GlcNAc_2$)	Variable	Highly variable
Flippase	Rft1?	PglK	Unknown
Oligosaccharyl transferase	OT multimeric	PglB monomeric	AglB possibly monomeric (?)
Protein recognition sequence	N-X-S/T	D/E-X-N-X-S/T	N-X-S/T
Timing of glycan transfer	Cotranslational, Posttranslational (?)	Posttranslational	Posttranslational
Glycan Processing	ER, Golgi	None	Unclear

Across all three kingdoms, the biosynthesis of *N*-linked glycans involves the buildup of an oligosaccharide onto a polyisoprenyl carrier by a series of glycosyltransferases using a set of discrete activated glycosyl donors. In eukaryotes, these donors are both nucleotide sugars and dolichyl-linked sugars, depending on the location of the glycosyltransferases relative to the ER membrane. In contrast, bacteria have been shown to exclusively rely on UDP-sugars for glycan donor biosynthesis. The glycosyl donors in archaea have not yet been clearly defined, though

the presence of both dolichylphosphate and dolichyldiphosphate suggest that perhaps archaea may use a combination of nucleotide- and dolichyl-activated sugars similar to eukaryotes. In all kingdoms, the polyisoprenyl-linked oligosaccharide is flipped across the membrane prior to its *en bloc* transfer to protein. The mechanism of flipping between the two faces of the membrane remains elusive. Although genetic screens originally implicated Rft1 in translocation of the heptasaccharide intermediate in *S. cerevisiae*, recent studies have found that this protein does not appear to be involved after all. The ABC transporter PglK has been identified as the flippase protein in *C. jejuni*; however, biochemical confirmation of this activity has not yet been reported. Interestingly, analysis of complete archaeal genome sequences using either Rft1 or PglK as a search model does not result in any putative flippase candidates, even though it is clear that the archaeal OTase catalyzes protein transfer on the external side of the cell membrane. Further work is needed to fully understand this interesting phenomenon.

In all three kingdoms, the action of an oligosaccharyl transferase is responsible for transferring the assembled oligosaccharide to asparagine side chains within acceptor proteins. Although the role of each of the subunits within the eukaryotic OT complex is still unresolved, recent evidence has implicated the Stt3 subunit directly in catalysis. PglB, the Stt3 homolog in *C jejuni*, has been shown sufficient to carry out this reaction in vitro, and preliminary studies involving the archaeal Stt3 homolog in *P. furiosus* suggest that it may function alone as well, though further evidence is still required. The simplicity of the prokaryotic OTases in contrast to the eukaryotic OT complex offers a powerful model system to begin to examine this reaction in greater depth. One of the major unresolved questions in this area involves the catalytic mechanism, namely the steps by which the asparagine side chain amide is activated by OT to affect oligosaccharide transfer. Interestingly, it appears that nearly all *N*-linked glycan structures

reported to date contain an acetylated amine at the C2" position of the sugar proximal to the modified asparagine residue; the one exception being the *N*-linked pentasaccharide of *H. halobium*, in which a glucose was found to be the linking sugar (Figure 1-13A). This observation supports earlier predictions that the 2-acetamido group may play an important role in the enzyme mechansim.[151]

As described in this chapter, the *N*-linked glycans observed in nature display a great deal of structural diversity. It appears that one of the major differences between the three kingdoms involves the origin of this diversity; specifically, whether it is introduced during the assembly process through the use of a variety of glycosyl donors, such as in bacteria, or after assembly and transfer to protein by the action of processing enzymes, as in the case of eukaryotes. Nearly all eukaryotes assemble a tetradecasaccharide ($Glc_3Man_9GlcNAc_2$) that is transferred to acceptor proteins by OT; however, this core structure is then modified by a series of glycosidases and glycosyltransferases in both the ER and Golgi to result in the great variety of eukaryotic *N*-linked glycans observed in nature. The conservation of this glycan core during the assembly process enables eukaryotes to engage the calnexin/calreticulin pathway for quality control of protein folding immediately after glycosylation, while still allowing for diversification of the glycan through the later steps. In contrast, the *N*-linked glycan structures identified on all *C. jejuni* glycoproteins to date are identical, indicating that the organism does not modify these oligosaccharides after protein transfer. The involvement of *N*-linked glycans in virulence may provide an explanation for the uniformity of bacterial glycoproteins, although the details about the roles of bacterial *N*-linked glycans in pathogenesis are still unresolved. It will be interesting to learn whether archaea reflect the eukaryotic or bacterial pathways in this regard, or if they derive glycan diversity in a completely unique manner.

In summary, the *N*-linked glycosylation pathways found in eukaryotes, bacteria and archaea share a great deal in common, but also exhibit unique differences specific to each kingdom. Future work in this field will focus on a variety of fronts, including gaining a deeper understanding of the organization of the eukaryotic OT complex, the interactions of bacterial glycoproteins with the mammalian immune system, the biochemical details of archaeal glycan assembly and transfer, and the exploitation of viral glycan profiles for disease treatment. Overall, it is an exciting time in glycobiology, as the next few years promise to continue to highlight the important role that *N*-linked glycosylation plays across all kingdoms of life.

Thesis focus and outline

The overall goal of the work described in this thesis is to gain a deeper insight into the process of *N*-linked glycosylation in prokaryotes. In Chapter 2, efforts to identify a suitable OTase homolog for in depth biochemical characterization in archaea are described. These studies determined that the OTases from *M. voltae* and *M. maripaludis* exhibited the highest levels of heterologous expression in *E. coli*. In order to generate a suitable substrate to probe the function of these enzymes, we required a robust synthesis of the nucleotide sugar UDP-GlcNAc(3NAc)A. The chemoenzymatic synthesis of this unique UDP-sugar harnesses the enzymes WbpB, WbpE and WbpD from the opportunistic pathogen *Pseudomonas aeruginosa* PAO1 and is outlined in Chapter 3. In Chapter 4, the crystallization and structure determination of the PLP-dependent aminotransferase WbpE is described as the first step toward gaining insight into the novel NAD^+-recycling pathway involving WbpB and WbpE. Chapter 5 summarizes attempts toward the identification of a functional UDP-GlcNAc(3NAc)A glycosyltransferase to generate our desired Dol-PP-disaccharide for study of the *M. voltae* and

M. maripaludis OTase. Finally, Chapter 6 describes efforts towards the development of an inhibitor of PglB, the bacterial OTase in *C. jejuni*.

References

1. Varki, A.; Cummings, R.D.; Freeze, H.H.; Stanley, P.; Bertozzi, C.R.; Hart, G.W.; Etzler, M.E. *Essentials of Glycobiology*; 2nd Ed.; Cold Spring Harbor Laboratory Press: New York, 2008.
2. Spiro, R.G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002**, *12*, 43R-56R.
3. Zarschler, K.; Janesch, B.; Pabst, M.; Altmann, F.; Messner, P.; Schäffer, C. Protein tyrosine *O*-glycosylation-A rather unexplored prokaryotic glycosylation system. *Glycobiology* **2010**, *20*, 787-798.
4. Mehta, D.P.; Etchinson, J.R.; Wu, R.; Freeze, H.H. UDP-GlcNAc:ser-protein *N*-acetylglucosamine-1-phosphotransferase from *Dictyostelium discoideum* recognizes serine-containing peptides and eukaryotic cysteine proteinases. *J. Biol. Chem.* **1997**, *272*, 28638-28645.
5. Haynes, P.A. Phosphoglycosylation: A new structural class of glycosylation? *Glycobiology* **1998**, *8*, 1-5.
6. Doucey, M.-A.; Hess, D.; Cacan, R.; Hofsteenge, J. Protein C-mannosylation is enzyme-catalyzed and uses dolicyl-phosphate-mannose as a precursor. *Mol. Biol. Cell.* **1998**, *9*, 291-300.
7. Nothaft, H.; Szymanski, C.M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **2010**, *8*, 765-778.
8. Calo, D.; Kaminski, L.; Eichler, J. Protein glycosylation in archaea: sweet and extreme. *Glycobiology* **2010**, *20*, 1065-1076.
9. Lehle, L.; Strahl, S.; Tanner, W. Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 6802-6818.
10. Heller, L.; Orlean, P.; Adair, W.L., Jr. *Saccharomyces cerevisiae* sec59 cells are deficient in dolichol kinase activity. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7013-7016.
11. Shiridas, P.; Waechter, C.J. Human dolichol kinase, a polytopic endoplasmic reticulum membrane protein with a cytoplasmically oriented CTP-binding site. *J. Biol. Chem.* **2006**, *281*, 31696-31704.
12. Schutzbach, J.S.; Jensen, J.W. Bilayer membrane destabilization induced by dolichylphosphate. *Chem Phys Lipids* **1989**, *51*, 213-218.
13. Chojnacki, T.; Dallner, G. The biological role of dolichol. *Biochem. J.* **1988**, *251*, 1-9.
14. Schenk, B.; Fernandez, F.; Waechter, C.J. The in(side) and outs(ide) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology* **2001**, *11*, 61R-70R.
15. Lehrman, M.A. Biosynthesis of N-acetylglucosamine-P-P-dolichol, the committed step of asparagine-linked oligosaccharide assembly. *Glycobiology* **1991**, *1*, 553-562.
16. Bickel, T.; Lehle, L.; Schwarz, M.; Aebi, M.; Jakob, C.A. Biosynthesis of lipid-linked oligosaccharides in *Saccharomyces cerevisiae*: Alg13p and Alg14p form a complex required for the formation of GlcNAc2-PP-Dolichol. *J. Biol. Chem.* **2005**, *280*, 34500-34506.
17. Gao, X.-D.; Tachikawa, H.; Sato, T.; Jigami, Y.; Dean, N. Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-*N*-

acetylglucosamine transferase required for the second step of N-linked glycosylation. *J. Biol. Chem.* **2005**, *280*, 36254-36262.

18. Gao, X.-D.; Moriyama, S.; Miura, N.; Dean, N.; Nishimura, S.-I. Interactions between the C-Termini of Alg13 and Alg14 mediates formation of the active UDP-*N*-acetylglucosamine transferase complex. *J. Biol. Chem.* **2008**, *283*, 32534-32541.
19. Wang, X.; Weldeghorghis, T.; Zhang, G.; Imperiali, B.; Prestegard, J.H. Solution structure of Alg13: the sugar donor subunit of a yeast *N*-acetylglucosamine transferase. *Structure* **2008**, *16*, 965-975.
20. Couto, J.R.; Huffaker, T.C.; Robbins, P.W. Cloning and expression in *Escherichia coli* of a yeast mannosyltransferase from the asparagine-linked glycosylation pathway. *J. Biol. Chem.* **1984**, *259*, 378-382.
21. Watt, G.M.; Revers, L.; Webberley, M.C.; Wilson, I.B.H.; Flitsch, S.L. Efficient enzymatic synthesis of the core trisaccharide of *N*-glycans with a recombinant β-mannosyltransferase. *Angew Chem Int Ed Engl* **1997**, *36*, 2354-2356.
22. Huffaker, T.C.; Robbins, P.W. Yeast mutants deficient in protein glycosylation. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 7466-7470.
23. Yamakazi, H.; Shiraishi, N.; Takauchi, K.; Ohnishi, Y.; Horinouchi, S. Characterization of ALG2 encoding a mannosyltransferase in the zygomycete fungus *Rhizomucor pusillus*. *Gene* **1998**, *221*, 179-184.
24. Cipollo, J.F.; Trimble, R.B.; Chi, J.H.; Dean, N. The yeast *ALG11* gene specifies addition of the terminal α1,2-Man to the $Man_5GlcNAc_2$-PP-dolichol *N*-glycosylation intermediate formed on the cytosolic side of the endoplasmic reticulum. *J. Biol. Chem.* **2001**, *276*, 21828-21840.
25. O'Reilly, M.K.; Zhang, G.; Imperiali, B. In vitro evidence for the dual function of Alg2 and Alg11: essential mannosyltransferases in *N*-linked glycoprotein biosynthesis. *Biochemistry* **2006**, *45*, 9593-9603.
26. Gao, X.-D.; Nishikawa, A.; Dean, N. Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum. *Glycobiology* **2004**, *14*, 559-570.
27. Noffz, C.; Keppler-Ross, S.; Dean, N. Hetero-oligomeric interactions between early glycosyltransferases of the dolichol cycle. *Glycobiology* **2009**, *19*, 472-478.
28. Helenius, J.; Ng, D.T.W.; Marolda, C.; Walter, P.; Valvano, M.A.; Aebi, M. Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **2002**, *415*, 447-450.
29. Frank, C.G.; Sanyal, S.; Rush, J.S.; Waechter, C.J.; Menon, A.K. Does Rft1 flip an N-linked glycan precursor? *Nature* **2008**, *454*, E3-E5.
30. Sanyal, S.; Menon, A.K. Specific transbilayer translocation of dolichol-linked oligosaccharides by an endoplasmic reticulum flippase. *Proc. Natl. Acad. Sci. USA* **2008**, *106*, 767-772.
31. Sanyal, S.; Menon, A.K. Flipping lipids: why an' what's the reason for? *ACS Chem. Biol.* **2009**, *4*, 895-909.
32. Rush, J.S.; Gao, N.; Lehrman, M.A.; Matveev, S.; Waechter, C.J. Suppression of Rft1 expression does not impair the transbilayer movement of Man5GlcNAc2-PP-Dolichol in sealed microsomes from yeast. *J. Biol. Chem.* **2009**, *284*, 19835-19842.
33. Burda, P.; Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta* **1999**, *1426*, 239-257.

34. Frank, C.G.; Aebi, M. *ALG9* mannosyltransferase is involved in two different steps of lipid-linked oligosaccharide biosynthesis. *Glycobiology* **2005**, *15*, 1156-1163.
35. Maeda, Y.; Kinoshita, T. Dolichol-phosphate mannose synthease: Structure, function and regulation. *Biochim. Biophys. Acta* **2008**, *1780*, 861-868.
36. Schutzbach, J.S.; Zimmerman, J.W.; Forsee, W.T. The purification and characterization of recombinant yeast dolichyl-phosphate-mannose synthase. Site-directed mutagenesis of the putative dolichol recognition sequence. *J. Biol. Chem.* **1993**, *268*, 24190-24196.
37. Sanyal, S.; Menon, A.K. Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11289-11294.
38. Heesen, S.t.; Lehle, L.; Weissmann, A.; Aebi, M. Isolation of the ALG5 locus encoding the UDP-glucose:dolichyl-phosphate glucosyltransferase from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **1994**, *224*, 71-79.
39. Guha-Niyogi, A.; Sullivan, D.R.; Turco, S.J. Glycoconjugate structures of parasitic protozoa. *Glycobiology* **2001**, *11*, 45R-59R.
40. Parodi, A.J. N-Glycosylation in trypanosomatid protozoa. *Glycobiology* **1993**, *3*, 193-199.
41. Samuelson, J.; Banerjee, S.; Magnelli, P.; Cui, J.; Kelleher, D.J.; Gilmore, R.; Robbins, P.W. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1548-1553.
42. Berhe, S.; Gerold, P.; Kedees, M.H.; Holder, A.A.; Schwarz, R.T. *Plasmodium falciparum*: merozoite surface proteins 1 and 2 are not posttranslationally modified by classical *N*- or *O*-glycans. *Exp. Parasitol.* **2000**, *94*, 194-197.
43. Dempski Robert, E., Jr.; Imperiali, B. Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr Opin Chem Biol* **2002**, *6*, 844-850.
44. Kelleher, D.J.; Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **2006**, *16*, 47R-62R.
45. Wilson, C.M.; Roebuck, Q.; High, S. Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9534-9539.
46. Ruiz-Canada, C.; Kelleher, D.J.; Gilmore, R. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell* **2009**, *136*, 272-283.
47. Li, H.; Chavan, M.; Schindelin, H.; Lennarz, W.J.; Li, H. Structure of the oligosaccharyl transferase complex at 12 Å resolution. *Structure* **2008**, *16*, 432-440.
48. Pathak, R.; Hendrickson, T.L.; Imperiali, B. Sulfhydryl modification of the yeast Wbp1p inhibits oligosaccharyl transferase activity. *Biochemistry* **1995**, *34*, 4179-4185.
49. Schulz, B.L.; Stirnimann, C.U.; Grimshaw, J.P.A.; Brozzo, M.S.; Fristsch, F.; Mohorko, E.; Capitani, G.; Glockshuber, R.; Grütter, M.G.; Aebi, M. Oxidoreductase activity of oligotransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11061-11066.
50. Glover, K.J.; Weerapana, E.; Numao, S.; Imperiali, B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **2005**, *12*, 1311-1315.

51. Parsaie Nasab, F.; Schulz, B.L.; Gamarro, F.; Parodi, A.J.; Aebi, M. All in one: *Leishmania major* STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **2008**, *19*, 3758-3768.
52. Hese, K.; Otto, C.; Routier, F.o.H.; Lehle, L. The yeast oligosaccharyltransferase complex can be replaced by STT3 from *Leishmania major*. *Glycobiology* **2009**, *19*, 160-171.
53. Manthri, S.; Güther, M.L.S.; Izquierdo, L.; Acosta-Serrano, A.; Ferguson, M.A.J. Deletion of the TbALG3 gene demonstrates site-specific *N*-glycosylation and *N*-glycan processing in *Trypanosoma brucei*. *Glycobiology* **2008**, *18*, 367-383.
54. Petrescu, A.-J.; Millac, A.-L.; Petrescu, S.M.; Dwek, R.A.; Wormald, M.R. Statistical analysis of the protein environment of *N*-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **2004**, *14*, 103-114.
55. Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **1983**, *209*, 331-336.
56. Gavel, Y.; von Heijne, G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* **1990**, *3*.
57. Nilsson, I.M.; von Heijne, G. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* **1993**, *268*, 5798-5801.
58. O'Connor, S.E.; Imperiali, B. Conformational switching by asparagine-linked glycosylation. *J. Am. Chem. Soc.* **1997**, *119*, 2295-2296.
59. Imperiali, B.; Shannon, K.L.; Unno, M.; Rickert, K.W. A mechanistic proposal for asparagine-linked glycosylation. *J. Am. Chem. Soc.* **1992**, *114*, 7944-7945.
60. Freeze, H.H. Genetic defects in the human glycome. *Nat. Rev. Genet.* **2006**, *7*, 537-551.
61. Moreman, K.M.; Molinari, M. *N*-linked glycan recognition and processing: the molecular basis of endoplasmic reticulum quality control. *Curr. Opin. Struct. Biol.* **2006**, *16*, 592-599.
62. Aebi, M.; Bernasconi, R.; Clerc, S.; Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends Biochem. Sci.* **2009**, *35*, 74-82.
63. Parodi, A.J. Protein glucosylation and its role in protein folding. *Annu. Rev. Biochem.* **2000**, *69*, 69-93.
64. Molinari, M. N-glycan structure dictates extension of protein folding or onset of disposal. *Nat Chem Biol* **2007**, *3*, 313-320.
65. Vembar, S.S.; Brodsky, J.L. One step at a time: endoplasmic reticulum-associated degradation. *Nat Rev Mol Cell Biol* **2008**, *9*, 944-957.
66. Dennis, J.W.; Lau, K.S.; Demetriou, M.; Nabi, I.R. Adaptive regulation at the cell surface by N-glycosylation. *Traffic* **2009**, *10*, 1569-1578.
67. Wyss, D.; Wagner, G. The structural role of sugars in glycoproteins. *Curr. Opin. Biotechnol.* **1996**, *7*, 409-416.
68. Kundra, R.; Kornfeld, S. Asparagine-linked oligosaccharides protect Lamp-1 and Lamp-2 from intracellular proteolysis. *J. Biol. Chem.* **1999**, *274*, 31039-31046.
69. Lau, K.S.; Dennis, J.W. N-glycans in cancer progression. *Glycobiology* **2008**, *18*, 750-760.

70. Arnold, J.N.; Saldova, R.; Hamid, U.M.A.; Rudd, P.M. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **2008**, *8*, 3284-3293.
71. Schmidt, M.A.; Riley, L.W.; Benz, I. Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* **2003**, *11*, 554-561.
72. Messner, P. Prokaryotic glycoproteins: unexplored but important. *J. Bacteriol.* **2004**, *186*, 2517-2519.
73. Hitchen, P.G.; Dell, A. Bacterial glycoproteomics. *Microbiol.* **2006**, *152*, 1575-1580.
74. Szymanski, C.M.; Yao, R.; Ewing, C.P.; Trust, T.J.; Guerry, P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **1999**, *32*, 1022-1030.
75. Szymanski, C.M.; Burr, D.H.; Guerry, P. *Campylobacter* protein glycosylation affects host cell interactions. *Infect. Immun.* **2002**, *70*, 2242-2244.
76. Linton, D.; Allan, E.; Karlyshev, A.V.; Cronshaw, A.D.; Wren, B.W. Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in Campylobacter jejuni. *Mol. Microbiol.* **2002**, *43*, 497-508.
77. Young, N.M.; Brisson, J.-R.; Kelly, J.; Watson, D.C.; Tessier, L.; Lanthier, P.H.; Jarrell, H.C.; Cadotte, N.; St. Michael, F.; Aberg, E.; Szymanski, C.M. Structure of the *N*-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **2002**, *277*, 42530-42539.
78. Jones, M.B.; Rosenberg, J.N.; Betenbaugh, M.J.; Krag, S.S. Structure and synthesis of polyisoprenoids used in *N*-glycosylation across the three domains of life. *Biochim. Biophys. Acta* **2009**, *1790*, 485-494.
79. Thibodeaux, C.J.; Melançon, C.E.; Liu, H.-w. Unusual sugar biosynthesis and natural product glycodiversification. *Nature* **2007**, *446*, 1008-1016.
80. Holden, H.M.; Cook, P.D.; Thoden, J.B. Biosynthetic enzymes of unusual microbial sugars. *Curr. Opin. Struct. Biol.* **2010**, *20*, 543-550.
81. Sharon, N. Celebrating the golden anniversary of the discovery of bacillosamine, the diamino sugar of a *Bacillus*. *Glycobiology* **2007**, *17*, 1150-1155.
82. Kocharova, N.A.; Perepelov, A.V.; Zatonsky, G.V.; Shashkov, A.S.; Knirel, Y.A.; Jansson, P.-E.; Weintraub, A. Structural studies of the O-specific polysaccharide of *Vibrio cholerae* O8 using solvolysis with triflic acid. *Carb. Res.* **2001**, *330*, 83-92.
83. Stimson, E.; Virji, M.; Makepeace, K.; Dell, A.; Morris, H.R.; Payne, G.; Saunders, J.R.; Jennings, M.P.; Barker, S.; Panico, M.; Blench, I.; Moxon, E.R. Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **1995**, *17*, 1201-1214.
84. Hanuszkiewicz, A.; Kaczyński, Z.; Lindner, B.; Goldmann, T.; Vollmer, E.; Debarry, J.; Heine, H.; Holst, O. Structural analysis of the capsular polysaccharide from *Acinetobacter lwoffii* F78. *Eur. J. Org. Chem.* **2008**, *2008*, 6183-6188.
85. Schoenhofen, I.C.; McNally, D.J.; Vinogradov, E.; Whitfield, D.; Young, N.M.; Dick, S.; Wakarchuk, W.W.; Brisson, J.-R.; Logan, S.M. Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: Enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *J. Biol. Chem.* **2006**, *281*, 723-732.

86. Olivier, N.B.; Chen, M.M.; Behr, J.R.; Imperiali, B. In vitro biosynthesis of UDP-*N*,*N*'-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry* **2006**, *45*, 13659-13669.
87. Glover, K.J.; Weerapana, E.; Chen, M.M.; Imperiali, B. Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* *N*-linked glycosylation pathway. *Biochemistry* **2006**, *45*, 5343-5350.
88. Weerapana, E.; Glover, K.J.; Chen, M.M.; Imperiali, B. Investigating bacterial *N*-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **2005**, *127*, 13766-13767.
89. Glover, K.J.; Weerapana, E.; Imperiali, B. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14255-14259.
90. Chen, M.M.; Weerapana, E.; Ciepichal, E.; Stupak, J.; Reid, C.W.; Swiezewska, E.; Imperiali, B. Polyisoprenol specificity in the *Campylobacter jejuni* *N*-linked glycosylation pathway. *Biochemistry* **2007**, *46*, 14342-14348.
91. Troutman, J.M.; Imperiali, B. *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. *Biochemistry* **2009**, *48*, 2807-2816.
92. Wacker, M.; Linton, D.; Hitchen, P.G.; Nita-Lazar, M.; Haslam, S.M.; North, S.J.; Panico, M.; Morris, H.R.; Dell, A.; Wren, B.W.; Aebi, M. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **2002**, *298*, 1790-1793.
93. Baar, C.; Eppinger, M.; Raddatz, G.; Simon, J.r.; Lanz, C.; Klimmek, O.; Nandakumar, R.; Gross, R.; Rosinus, A.; Keller, H.; Jagtap, P.; Linke, B.; Meyer, F.; Lederer, H.; Schuster, S.C. Complete genome sequence and analysis of *Wolinella succinogenes*. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11690-11695.
94. Santos-Silva, T.; Dias, J.o.M.; Dolla, A.; Durand, M.-C.; GonÁalves, L.L.; Lampreia, J.; Moura, I.; Rom„o, M.J.o. Crystal structure of the 16 heme cytochrome from *Desulfovibrio gigas*: a glycosylated protein in a sulphate-reducing bacterium. *J. Mol. Biol.* **2007**, *370*, 659-673.
95. Jervis, A.J.; Langdon, R.; Hitchen, P.; Lawson, A.J.; Wood, A.; Fothergill, J.L.; Morris, H.R.; Dell, A.; Wren, B.; Linton, D. Characterization of N-Linked Protein Glycosylation in Helicobacter pullorum. *J. Bacteriol.* **2010**, *192*, 5228-5236.
96. Power, P.M.; Roddam, L.F.; Rutter, K.; Fitzpatrick, S.Z.; Srikhanta, Y.N.; Jennings, M.P. Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. *Mol. Microbiol.* **2003**, *49*, 833-847.
97. Vik, Å.; Aas, F.E.; Anonsen, J.H.; Bilsborough, S.; Schneider, A.; Egge-Jacobsen, W.; Koomey, M. Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4447-4452.
98. Faridmoayer, A.; Fentabil, M.A.; Mills, D.C.; Klassen, J.S.; Feldman, M.F. Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. *J. Bacteriol.* **2007**, *189*, 8088-8098.
99. Maita, N.; Nyirenda, J.; Igura, M.; Kamishikiryo, J.; Kohda, D. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J. Biol. Chem.* **2010**, *285*, 4941-4950.

100. Kowarik, M.; Numao, S.; Feldman, M.F.; Schulz, B.L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M. *N*-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **2006**, *314*, 1148-1150.
101. Nita-Lazar, M.; Wacker, M.; Schegg, B.; Amber, S.; Aebi, M. The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **2005**, *15*, 361-367.
102. Chen, M.M.; Glover, K.J.; Imperiali, B. From peptide to protein: comparative analysis of the substrate specificity of *N*-linked glycosylation in *C. jejuni*. *Biochemistry* **2007**, *46*, 5579-5585.
103. Rangarajan, E.S.; Bhatia, S.; Watson, D.C.; Munger, C.; Cygler, M.; Matte, A.; Young, N.M. Structural context for protein *N*-glycosylation in bacteria: the structure of PEB3, an adhesin from *Campylobacter jejuni*. *Protein Sci.* **2007**, *16*, 990-995.
104. Slynko, V.; Schubert, M.; Numao, S.; Kowarik, M.; Aebi, M.; Allain, F.H.T. NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J. Am. Chem. Soc.* **2009**, *131*, 1274-1281.
105. Feldman, M.F.; Wacker, M.; Hernandez, M.; Hitchen, P.G.; Marolda, C.L.; Kowarik, M.; Morris, H.R.; Dell, A.; Valvano, M.A.; Aebi, M. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3016-3021.
106. Hendrixson, D.R.; DiRita, V.J. Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract. *Mol. Microbiol.* **2004**, *52*, 471-484.
107. Van Sorge, N.M.; Bleumink, N.M.C.; Van Vliet, S.J.; Saeland, E.; Van Der Pol, W.L.; Van Kooyk, Y.; Van Putten, J.P.M. *N*-glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL. *Cell. Microbiol.* **2009**, *11*, 1768-1781.
108. Larsen, J.C.; Szymanski, C.; Guerry, P. N-linked protein glycosylation is required for full competence in *Campylobacter jejuni* 81-176. *J. Bacteriol.* **2004**, *186*, 6508-6514.
109. Nothaft, H.; Liu, X.; McNally, D.J.; Li, J.; Szymanski, C.M. Study of free oligosaccharides derived from the bacterial N-glycosylation pathway. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15019-15024.
110. Woese, C.R.; Fox, G.E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5088-5090.
111. Sprott, G.D. Structures of archaebacterial membrane lipids. *J. Bioenerg. Biomembr.* **1992**, *24*, 555-566.
112. Chaban, B.; Ng, S.Y.M.; Jarrell, K.F. Archaeal habitats--from the extreme to the ordinary. *Can. J. Microbiol.* **2006**, *52*, 73-116.
113. Mescher, M.F.; Strominger, J.L. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *J. Biol. Chem.* **1976**, *251*, 2005-2014.
114. Wieland, F.; Heitzer, R.; Schaefer, W. Asparaginylglucose: Novel type of carbohydrate linkage. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 5470-5474.
115. Lechner, J.; Wieland, F. Structure and biosynthesis of prokaryotic glycoproteins. *Annu. Rev. Biochem.* **1989**, *58*, 173-194.
116. Wieland, F.; Paul, G.; Sumper, M. Halobacterial flagellins are sulfated glycoproteins. *J. Biol. Chem.* **1985**, *260*, 15180-15185.

117. Paul, G.; Lottspeich, F.; Wieland, F. Asparaginyl-N-acetylgalactosamine. Linkage unit of halobacterial glycosaminoglycan. *J. Biol. Chem.* **1986**, *261*, 1020-1024.
118. Kärcher, U.; Schröder, H.; Haslinger, E.; Allmaier, G.; Schreiner, R.; Wieland, F.; Haselbeck, A.; König, H. Primary structure of the heterosaccharide of the surface glycoprotein of *Methanothermus fervidus*. *J. Biol. Chem.* **1993**, *268*, 26821-26826.
119. Voisin, S.; Houliston, R.S.; Kelly, J.; Brisson, J.-R.; Watson, D.; Bardy, S.L.; Jarrell, K.F.; Logan, S.M. Identification and characterization of the unique *N*-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *J. Biol. Chem.* **2005**, *280*, 16586-16593.
120. Chaban, B.; Logan, S.M.; Kelly, J.F.; Jarrell, K.F. AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the *N*-Glycan in *Methanococcus voltae*. *J. Bacteriol.* **2009**, *191*, 187-195.
121. Kelly, J.; Logan, S.M.; Jarrell, K.F.; VanDyke, D.J.; Vinogradov, E.V. A novel *N*-linked flagellar glycan from *Methanococcus maripaludis*. *Carb. Res.* **2009**, *344*, 648-653.
122. Zähringer, U.; Moll, H.; Hettmann, T.; Knirel, Y.A.; Schäfer, G. Cytochrome $b_{558/566}$ from the archaeon *Sulfolobus acidocaldarius* has a unique Asn-linked highly branched hexasaccharide chain containing 6-sulfoquinovose. *Eur. J. Biochem.* **2000**, *267*, 4144-4149.
123. Yang, L.L.; Haug, A. Purification and partial characterization of a procaryotic glycoprotein from the plasma membrane of *Thermoplasma acidophilum*. *Biochim. Biophys. Acta* **1979**, *556*, 265-277.
124. Igura, M.; Maita, N.; Kamishikiryo, J.; Yamada, M.; Obita, T.; Maenaka, K.; Kohda, D. Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **2008**, *27*, 234-243.
125. Eichler, J.; Adams, M.W.W. Posttranslational protein modification in archaea. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 393-425.
126. Magidovich, H.; Eichler, J. Glycosyltransferases and oligosaccharyltransferases in archaea: putative components of the *N*-glycosylation pathway in the third domain of life. *FEMS Microbiol. Lett.* **2009**, *300*, 2005-2014.
127. Chaban, B.; Voisin, S.; Kelly, J.; Logan, S.M.; Jarrell, K.F. Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae N*-linked glycans: insight into *N*-linked glycosylation pathways in Archaea. *Mol. Microbiol.* **2006**, *61*, 259-268.
128. Mescher, M.F.; Hansen, U.; Strominger, J.L. Formation of lipid-linked sugar compounds in *Halobacterium salinarium*. Presumed intermediates in glycoprotein synthesis. *J. Biol. Chem.* **1976**, *251*, 7289-7294.
129. Lechner, J.; Wieland, F.; Sumper, M. Biosynthesis of sulfated saccharides N-glycosidically linked to the protein via glucose. Purification and identification of sulfated dolichyl monophosphoryl tetrasaccharides from halobacteria. *J. Biol. Chem.* **1985**, *260*, 860-866.
130. Zhu, B.C.R.; Drake, R.R.; Schweingruber, H.; Laine, R.A. Inhibition of Glycosylation by Amphomycin and Sugar Nucleotide Analogs PP36 and PP55 Indicates That *Haloferax volcanii* β-Glucosylates Both Glycoproteins and Glycolipids through Lipid-Linked Sugar Intermediates: Evidence for Three Novel Glycoproteins and a Novel Sulfated Dihexosyl-Archaeol Glycolpid. *Arch. Biochem. Biophys.* **1995**, *319*, 355-364.

131. Zhu, B.C.R.; Laine, R.A. Dolichyl-phosphomannose synthase from the Archae *Thermoplasma acidophilum*. *Glycobiology* **1996**, *6*, 811-816.
132. Shams-Eldin, H.; Chaban, B.; Niehus, S.; Schwarz, R.T.; Jarrell, K.F. Identification of the Archaeal *alg7* Gene Homolog (Encoding *N*-Acetylglucosamine-1-Phosphate Transferase) of the *N*-Linked Glycosylation System by Cross-Domain Complementation in *Saccharomyces cerevisiae*. *J. Bacteriol.* **2008**, *190*, 2217-2220.
133. Rush, J.S.; Alaimo, C.; Robbiani, R.; Wacker, M.; Waechter, C.J. A novel epimerase that converts GlclNAc-P-P-undecaprenol to GalNAc-P-P-undecaprenol in *Escherichia coli* O1570. *J. Biol. Chem.* **2010**, *285*, 1671-1680.
134. Jarrell, K.F.; Jones, G.M.; Nair, D.B. Biosynthesis and role of N-linked glycosylation in cell surface structures of archaea with a focus on flagella and S-layers. *Int. J. Microbiol.* **2010**, *2010*, 470138.
135. VanDyke, D.J.; Wu, J.; Logan, S.M.; Kelly, J.F.; Mizuno, S.; Aizawa, S.-I.; Jarrell, K.F. Identification of genes involves in the assembly and attachment of a novel flagellin *N*-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripaludis*. *Mol. Microbiol.* **2009**, *72*, 633-644.
136. Mescher, M.F.; Strominger, J.L. Glycosylation of the surface glycoprotein of *Halobacterium salinarium* via a cyclic pathway of lipid-linked intermediates. *FEBS Lett.* **1978**, *89*, 37-41.
137. Storm, D.R.; Strominger, J.L. Complex formation between bacitracin peptides and isoprenyl pyrophosphates. *J. Biol. Chem.* **1973**, *248*, 3940-3945.
138. VanDyke, D.J.; Wu, J.; Logan, S.M.; Kelly, J.F.; Mizuno, S.; Aizawa, S.-I.; Jarrell, K.F. Identification of genes involved in the assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripalidus*. *Mol. Microbiol.* **2009**, *72*, 633-644.
139. Abu-Qarn, M.; Eichler, J. An analysis of amino acid sequences surrounding archaeal glycoprotein sequons. *Archaea* **2007**, *2*, 73-81.
140. Valentine, D.L. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat. Rev. Microbiol.* **2007**, *5*, 316-323.
141. Madern, D.; Ebel, C.; Zaccai, G. Halophilic adaptation of enzymes. *Extremophiles* **2000**, *4*, 91-98.
142. Vieille, C.; Zeikus, G.J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1-43.
143. Luke, K.A.; Higgins, C.L.; Wittung-Stafshede, P. Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J.* **2007**, *274*, 4023-4033.
144. Jarrell, K.F.; McBride, M.J. The surprisingly diverse ways that prokaryotes move. *Nat. Rev. Microbiol.* **2008**, *6*, 446-476.
145. Vigerust, D.J.; Shepherd, V.L. Virus glycosylation: role in virulence and immune interactions. *Trends in Microbiology 15*, 211-218.
146. Keil, W.; Geyer, R.; Dabrowski, U.; Niemann, H.; Stirm, S.; Klenk, H.D. Carbohydrates of influenza virus. Structural elucidation of the individual glycans of the FPV hemagglutinin by two-dimensional 1H n.m.r. and methylation analysis. *EMBO J.* **1985**, *4*, 2711-2720.
147. Wang, C.-C.; Chen, J.-R.; Tseng, Y.-C.; Hsu, C.-H.; Hung, Y.-F.; Chen, S.-W.; Chen, C.-M.; Khoo, K.-H.; Cheng, T.-J.; Cheng, Y.-S.E.; Jan, J.-T.; Wu, C.-Y.; Ma, C.; Wong,

C.-H. Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18137-18142.

148. Doores, K.J.; Bonomelli, C.; Harvey, D.J.; Vasiljevic, S.; Dwek, R.A.; Burton, D.R.; Crispin, M.; Scanlan, C.N. Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13800-13805.
149. Korber, B.; Gaschen, B.; Yusim, K.; Thakallapally, R.; Kesmir, C.; Detours, V. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* **2001**, *58*, 19-42.
150. Calarese, D.A.; Lee, H.-K.; Huang, C.-Y.; Best, M.D.; Astronomo, R.D.; Stanfield, R.L.; Katinger, H.; Burton, D.R.; Wong, C.-H.; Wilson, I.A. Dissection of the carbohydrate specificity of the broadly neutralizing anti-HIV-1 antibody 2G12. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13372-13377.
151. Tai, V.W.F.; Imperiali, B. Substrate specificity of the glycosyl donor for oligosaccharyl transferase. *J. Org. Chem.* **2001**, *66*, 6217-6228.

Chapter 2 Identification and Characterization of an Archaeal Oligosaccharyl Transferase for In Depth Study

Introduction

Asparagine-linked glycosylation is a posttranslational protein modification that occurs across all domains life.[1] This process plays a critical role in many cellular events, including protein folding, stability and intracellular targeting in eukaryotes, flagellum assembly in archaea, and host pathogenicity in bacteria.[2-4] *N*-linked glycosylation entails the stepwise assembly of an oligosaccharide onto a membrane-bound polyisoprenyl-linked carrier, followed by the translocation of the glycan across the membrane and subsequent transfer to protein (Figure 2-1). The key enzyme of this pathway is oligosaccharyl transferase (OTase), which catalyzes the *en bloc* transfer of the oligosaccharide onto asparagine residues within nascent or fully folded proteins. This enzyme recognizes and modifies asparagine residues that reside within an N-X-S/T sequon, where X is any amino acid except proline.[5]

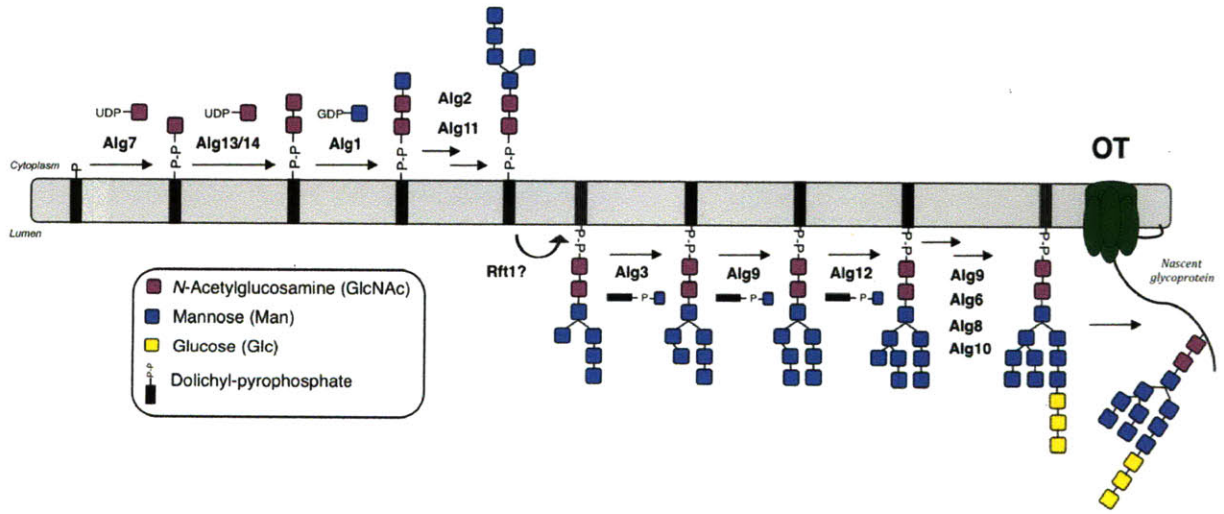


Figure 2-1: The pathway of *N*-linked glycosylation in *S. cerevisiae.*

In eukaryotes, the oligosaccharyl transferase (OT) is a multimeric protein complex localized in the membrane of the endoplasmic reticulum (ER). While a majority of the genetic

and biochemical characterization of OT has been achieved in *Saccharomyces cerevisiae*, the organization of the complex appears to be remarkably conserved across all eukaryotes, from yeast to man.[6] In *S. cerevisiae*, the OT complex is composed of at least eight membrane-bound subunits (Ost3/6, Ost4, Stt3, Ost2, Wbp1, Swp1, Ost5, and Ost1), five of which (Ost1, Ost2, Stt3, Swp1, and Wbp1) have been shown to be essential for survival (Figure 2-2A).[7,8] Although in depth biochemical examination has been hampered by difficulties involving overexpression, isolation, and handling of the complex, initial progress has been made toward identifying the role of each protein subunit.[8] The putative binding site of the dolichyl-linked oligosaccharide was assigned to Wbp1, as incubation of OT with Dol-PP-$GlcNAc_2$ was able to prevent alkyation of a specific cysteine located within the Wbp1 subunit.[9] The Ost3/Ost6 subunits have been recently shown to exhibit protein-dependent oxidoreductase activity in vitro, suggesting a direct role for OT in protein folding.[10] In recent years, evidence has begun to emerge implicating Stt3 as the catalytic subunit of the OT complex. Stt3 is the most highly conserved subunit of OT, as homologs of this protein are found in all eukaryotes and throughout the archaeal and bacterial kingdoms. Stt3 is composed of a large N-terminal domain containing 11-14 predicted transmembrane regions, as well as a C-terminal domain localized in the ER lumen that bears the highly conserved WWDXGX motif, which is found in all known Stt3 homologs in nature. Genetic mutations introduced into this motif have been shown to abolish glycosylation activity, strongly suggesting that these residues play a role in catalysis.[11] In addition, the recent cryo-EM structure of the OT complex at 12 Å suggests that Stt3, along with Wbp1 and Ost1, line a deep channel within the complex through which both the dolichyl-linked oligosaccharide and acceptor protein substrates are predicted to bind.[12]

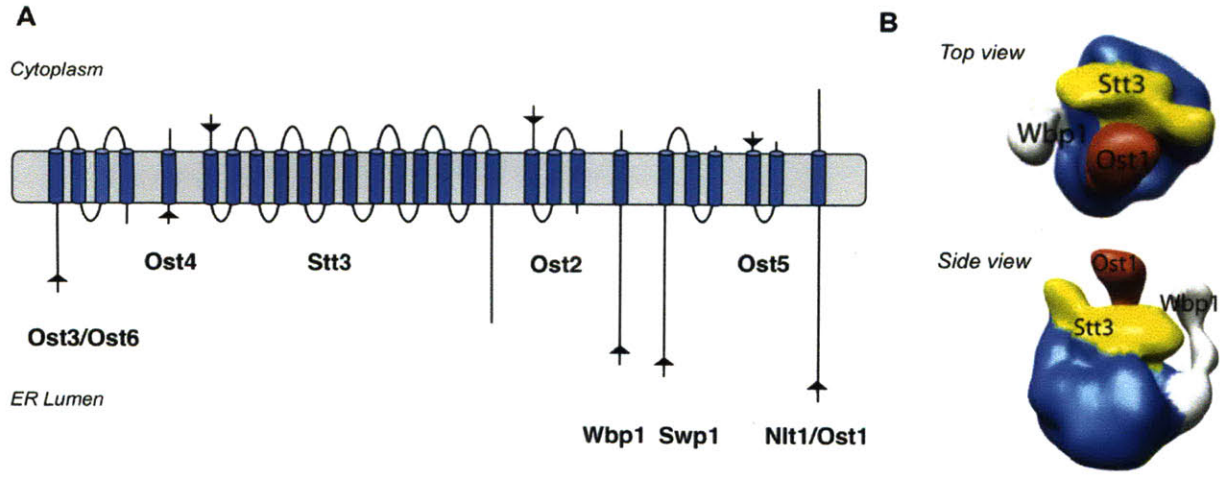


Figure 2-2: Subunits of the *S. cerevisiae* oligosaccharyl transferase complex. (A) Predicted topology of the individual subunits, where the arrows indicate the N-termini of the individual proteins; (B) Three-dimensional cryo-EM map of the OT complex, as reported by Li et al.[12]

In 1999, the first example of a general system of *N*-linked glycosylation in bacteria was reported in the human pathogen *Campylobacter jejuni*.[13] This pathway contains a number of similarities to the glycosylation pathway in *S. cerevisiae*, in that an oligosaccharide is assembled onto a polyisoprenyl-linked carrier and then flipped across the inner leaflet of the membrane before being transferred to acceptor proteins by an OTase (Figure 2-3). However, one of the most intriguing differences between these two pathways is that the *C. jejuni* OTase, PglB, is composed of just a single polypeptide. A comparison of PglB and Stt3 sequences show that the two proteins are homologous and share a similar topology, wherein the N-termini comprise a large hydrophobic domain containing 9-13 transmembrane helices, and the hydrophilic C-termini contain the conserved WWDXGX motif. In pivotal studies, it was demonstrated that PglB alone is sufficient to catalyze the transfer of the *C. jejuni* heptasaccharide onto fully folded acceptor proteins, providing a simpler platform for the in depth study of this reaction.[14,15] Further work showed PglB requires an extended glycosylation sequence, D-X-N-X-S/T, in

which an aspartate is required at the -2 position relative to asparagine, and that it demonstrates a relaxed oligosaccharide specificity in comparison to the yeast OT.[16-18] However, despite this progress, efforts to carry out in depth biochemical studies of PglB have been challenged by the extremely poor expression levels in heterologous systems (< 50 μg/L culture in *Escherichia coli*), as well as a rapid loss of activity upon purification of the enzyme from the membrane.[19] In order to begin to unravel the enzymatic mechanism of the oligosaccharide transfer reaction, a long-standing goal of the Imperiali laboratory, large quantities of a highly pure and active OTase homolog are required.

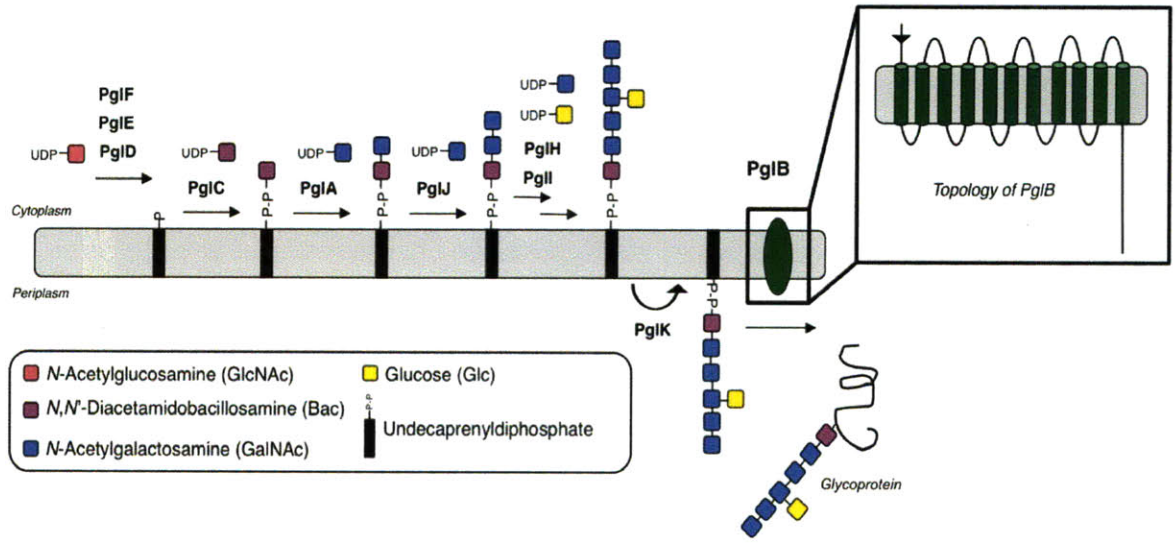


Figure 2-3: The *N*-linked glycosylation pathway of *C. jejuni*, highlighting the predicted membrane topology of PglB.

In an effort to obtain a tractable OTase with improved expression and stability for in depth examination, we have turned to the *N*-linked glycosylation pathway in archaea. Archaeal protein homologs have been long used to aid in analysis of protein structure and function due to increased expression in heterologous systems as well as enhanced stability compared with their

eukaryotic and bacterial counterparts.[20] The harsh environments inhabited by certain archaea, such as volcanic deep sea vents and acidic coal mines, confer characteristics to archaeal proteins that can be highly advantageous to researchers working in ambient conditions. As detailed in Chapter 1, the first archaeal glycoproteins were identified in the early 1970s in *Halobacterium salinarum*, an extreme halophile isolated from a highly saline lake in the Michigan Basin.[21] Although a number of archaeal glycoproteins have been characterized to date that display a wide variety of glycan structures, the biochemical details of the glycosylation pathways in archaea are scant. Recent studies suggest that *N*-linked glycosylation in archaea bears some similarities to both the bacterial and eukaryotic pathways. As in bacteria, the *N*-linked glycans of archaea vary tremendously based on species, and oligosaccharide transfer appears to be carried out by a single integral membrane protein.[22] However, unlike in bacteria, where *N*-linked glycosylation is considered a rare event, archaeal *N*-linked glycoproteins appear to be a prevalent feature of secretory proteins similar to eukaryotes, appearing in both S-layer and flagellar proteins.[3] In addition, the polyisoprenoid carrier appears to be dolichol, rather than the fully unsaturated undecaprenol characteristic of bacteria.[23] These findings make the pathway in archaea an attractive system for in depth biochemical characterization.

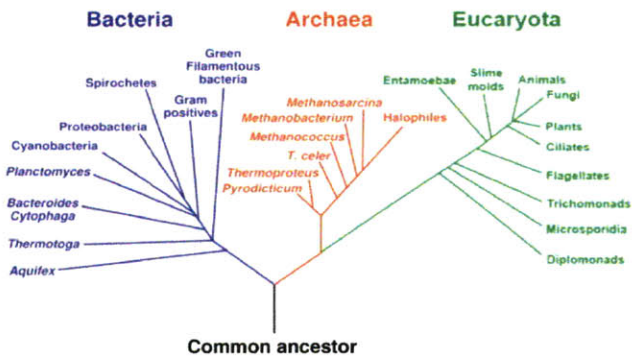


Figure 2-4: Phylogenetic tree depicting the three kingdoms of life: Eucaryota, Bacteria, and Archaea.[24]

In this chapter, efforts toward the identification and characterization of a suitable archaeal OTase are described. An initial search of available and completed archaeal genome sequences yielded a total of 25 archaeal Stt3 homologs, from which 5 candidates were selected for further analysis. After cloning these genes into suitable vectors, protein expression levels were compared in various *E. coli* cell lines and revealed that the Stt3 homologs from the methanogen *Methanococcus voltae* and the extreme halophile *Haloarcula marismortui* exhibited the highest overall protein expression yields, representing a nearly 100-fold improvement over PglB. Further work demonstrated that these proteins could be purified from the cell membrane; however, efforts to observe function of these proteins were unsuccessful due to the unavailibity of the natural oligosaccharide substrate of these enzymes. This work summarizes initial efforts to characterize an archaeal OTase, which is important both to provide a platform for the detailed understanding of the mechanism of oligosaccharide transfer and also as a first step towards the elucidation of the archaeal glycosylation pathway that appears to have elements in common with both the bacterial and eukaryotic systems.

Results and Discussion

Identification of Archaeal Stt3 Homologs

In order to identify potential archaeal OTase for study, iterative searches of completed archaeal genome sequences in the public databases (NCBI, GenBank) were performed using the *S. cerevisiae* Stt3 protein as the search model. A total of 25 distinct homologs were found with a local alignment score of 90 or greater, and all contained the WWDXGX motif common to

known oligosaccharyl transferases. A truncated alignment of 16 of these sequences is depicted in Figure 2-5, and a full alignment is presented in the Appendix (Figure 1).

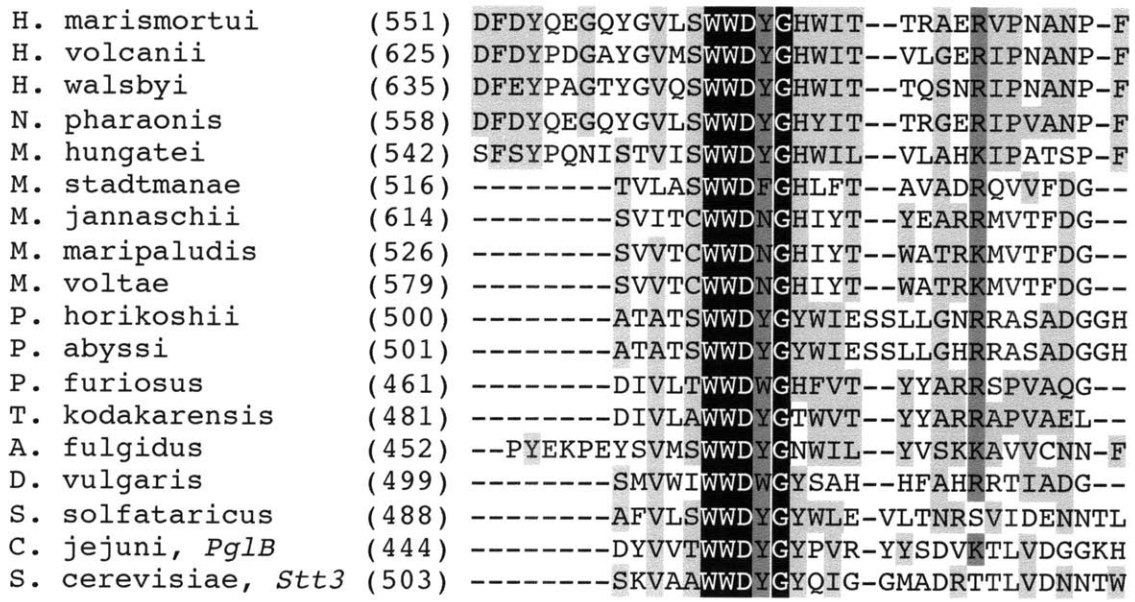


Figure 2-5: Sequence alignment of 16 of the 25 archaeal Stt3 homologs identified in this study, depicting the region surrounding the critical WWDXGX motif. Fully conserved residues are highlighted in black, while moderate conservation is indicated in gray. A full alignment of these homologs is presented in the Appendix (Figure 1).

Each of these protein sequences was further analyzed using the TMHMM server in order to predict membrane topology (ca.expasy.org/tools).[25] Similar to both Stt3 and PglB, these proteins are all predicted to contain a large N-terminal domain with between 9-12 transmembrane spanning helices, and a soluble C-terminal domain that bears the WWDXGX motif. Interestingly, the size of the archaeal Stt3 homologs varies widely, from short (*Archaeoglobus fulgidus*, 591 residues) to extremely long (*Thermoplasma volcanium*, 2076 residues). In addition, many of these predicted helices appeared to be more compact than Stt3 and PglB, and the loops between each membrane domain seemed to be shorter in length. In order to select a few OTase candidates for biochemical analysis, several factors such as protein

size, the thermal and/or salt environment of the parent species (perhaps an indicator of the stability of the proteins), phylogenetic diversity, and availability of genomic DNA were considered with the goal of assembling a small panel of proteins from a wide variety of archaeal strains. To this end, five proteins were selected from the following organisms for further study: the thermophiles *Methanocaldococcus jannaschii* and *Archaeoglobus fulgidus*, the mesophilic methanogens *Methanococcus voltae* and *Methanococcus maripaludis*, and the extreme halophile *Haloarcula marismortui*. The predicted membrane topology of these five proteins, as well as Stt3 and PglB, are compared in Figure 2-6.

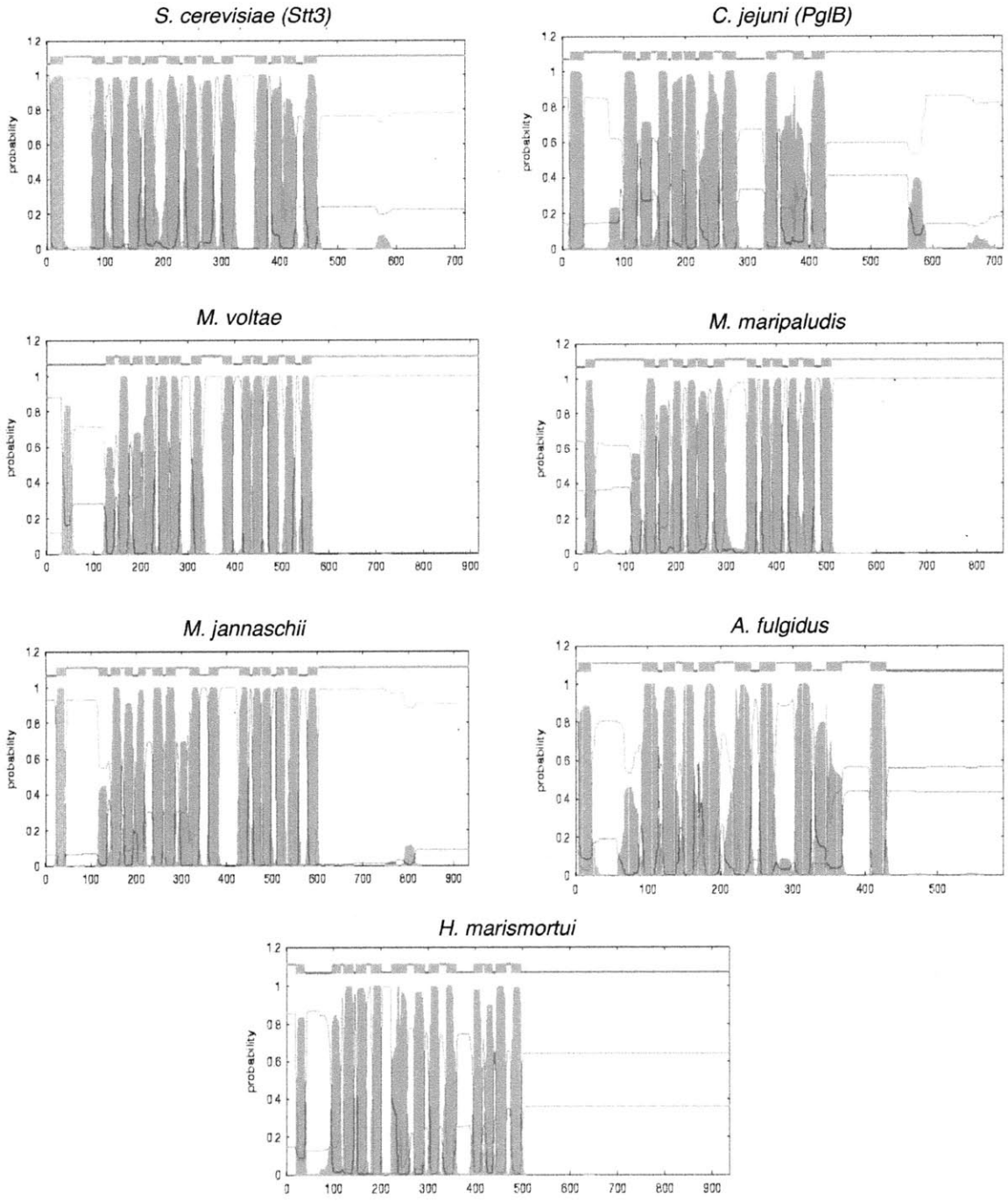


Figure 2-6: Predicted topology of the five selected archaeal Stt3 homologs in comparison with Stt3 (*S. cerevisiae*) and PglB (*C. jejuni*). Topology prediction was generated using the TMHMM server (ca.expasy.org/tools).[25]

Molecular Biology and Comparative Analysis of Protein Expression Levels

After genomic DNA from each of the five archaeal species was acquired, either through purchase (ATCC) or donation from either the Whitman (University of Georgia) and Jarrell (Queen's University) laboratories, the OTase genes were amplified using the polymerase chain reaction. All of these genes are large (3-4 kb) and exhibit a high GC content, which can manifest significant hydrophobic interactions in the genomic DNA that can make finding suitable PCR conditions a considerable challenge. Initial attempts at PCR were unsuccessful, but after screening a large number of conditions, including the use of additives such as glycerol, Triton X-100, and DMSO, and the manipulation of PCR cycle features, amplification of the desired genes was observed (Figure 2-7). The resulting PCR products were inserted into the pET24a(+) vector using standard molecular biology techniques to yield proteins with an N-terminal T7 tag for identification and a C-terminal His_6-tag for purification.

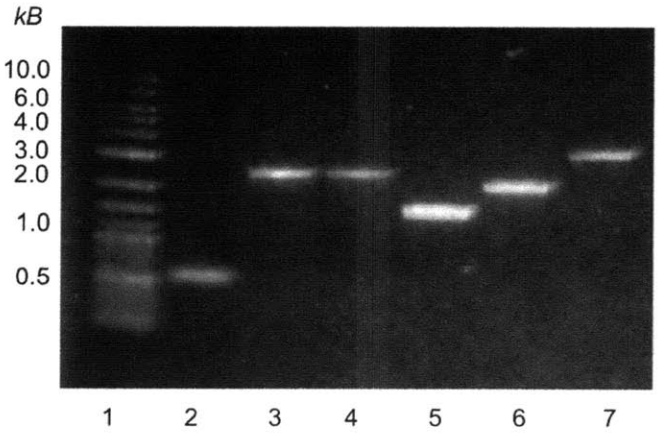


Figure 2-7: Agarose gel (0.8%) of the PCR products resulting from amplification of the *stt3* homologs. (1) MW Standard; (2) PCR positive control; (3) *M. voltae*; (4) *M. maripaludis*; (5) *A. fulgidus*; (6) *M. jannaschii*; (7) *H. marismortui*.

The resulting PCR products were then transformed into four different *E. coli* competent cell lines in order to screen for protein expression. Each of these cell lines possesses a distinct advantageous trait that may serve to improve the overall yield of the archaeal OT homologs; BL21(DE3) Gold cells exhibit a high transformation efficiency, while both BL21 CodonPlus RIL and RP cells contain extra copies of genes encoding for rare tRNAs. BL21(DE3) pLysS cells confer the transformation efficiency of the BL21(DE3) Gold cells and also lack genes for several proteases responsible for degrading nonnative proteins. After transformation of the archaeal OTase genes into each of these strains, along with the empty pET24a(+) vector as a control, all 20 cell lines were cultured and the proteins overexpressed using isopropyl β-D-1-thiogalactopyranoside (IPTG) as an inducer. Upon harvesting, the cells were lysed and the membrane fractions were isolated by centrifugation. The presence of the desired OTase was determined through Western blot analysis with the Anti-T7 antibody, which recognizes the N-terminal T7 tag on each protein.

	BL21 Gold	**BL21 RIL**	**BL21 RP**	**BL21 pLysS**
A. fulgidus	Strong	Strong	Weak	X
M. jannaschii	X	X	X	X
M. voltae	Strong	Very Strong	Strong	X
M. maripaludis	Very Strong	Strong	Strong	X
H. marismortui	Weak	Very Strong	Strong	X

Table 2-1: Summary of the protein expression trials of each archaeal OTase in all four *E. coli* competent cell lines. Levels of expression were determined by Western blot analysis, using an Anti-T7 antibody to detect the N-terminal T7 tag of each protein.

Table 2-1 summarizes the results of the archaeal OTase expression screen. As indicated, high expression levels of the *M. voltae*, *M. maripaludis*, *A. fulgidus*, and *H. marismortui* OTases were observed in three out of the four cell lines. Based on Western blot analysis, it was determined that the BL21 Gold cell line would be optimal for both the *A. fulgidus* and *M. maripalidus* proteins, while the BL21 RIL cell line was chosen for expression of the *M. voltae* and *H. marismortui* OTases. Interestingly, no expression of the *M. jannaschii* homolog was seen in any of the cell lines; further attempts to achieve protein expression, such as retransformation of the competent cell lines and varying the amount of time that the cultures were incubated, were unsuccessful. In addition, none of the four homologs was expressed in the BL21(DE3) pLysS cell line, though the reason for this result is unclear.

As a means to directly compare overall expression levels of each protein in the optimal *E. coli* strain, proteins from the best expressing cell line for each OTase were analyzed side by side on the same Western blot, using antibodies for both the N- and C-terminal tags (Figure 2-8). A sample of PglB at a known concentration was also included in a preliminary attempt to quantify total protein yield. These results indicated that the *M. voltae* and *H. marismortui* OTases exhibited the highest overall expression levels, with a nearly 100-fold improvement over PglB. The *M. maripaludis* and *A. fulgidus* OTases were also expressed in high yield, although these proteins showed a more modest improvement compared with PglB, representing an approximately 50- and 20-fold increase, respectively. Therefore, these four proteins were carried forward for further study.

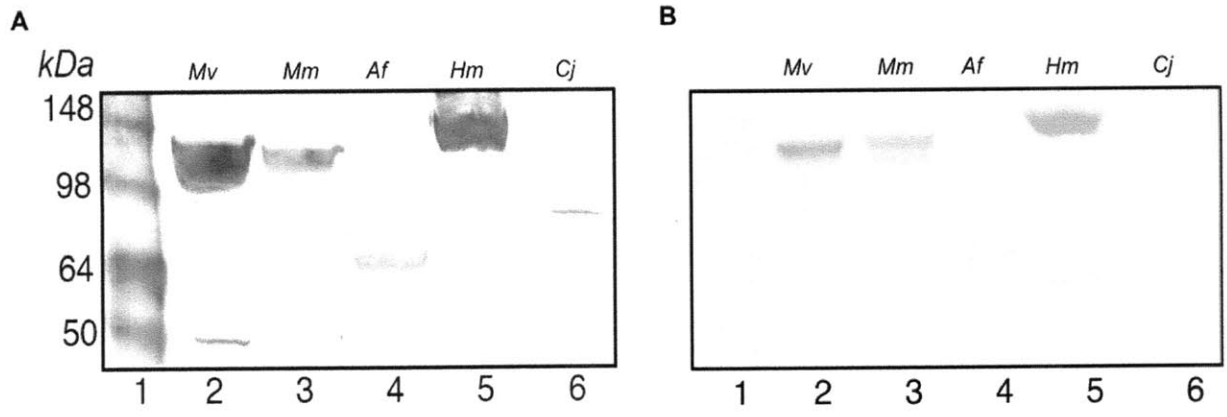


Figure 2-8: Comparison of the archaeal OTases expression by Western blot, in which each OTase is expressed in the optimal *E. coli* cell line as described. Each protein was prepared as a crude *E. coli* membrane fraction prior to analysis. (A) Anti-T7 and (B) Anti-His_4 Western blots, indicating the N- and C-terminal tags. (1) MW standard; (2) *M. voltae*; (3) *M. maripaludis*; (4) *A. fulgidus*; (5) *H. marismortui*; (6) *C. jejuni* (PglB).

Purification of Archaeal OTases

In order to obtain a quantitative measurement of overall expression levels, efforts were made to purify the archaeal OTases from the *E. coli* membrane using Ni-NTA chromatography. After overexpression of each OTase in the optimal *E. coli* cell line, the proteins were prepared as crude membrane fractions following standard procedures. The OTases were then incubated with Triton X-100 detergent (1% final concentration) for 8 hrs at 4 °C to disrupt the membranes. While Triton X-100 is often not an ideal detergent for various downstream applications, such as functional characterization or X-ray crystallography, it has been found to be a powerful agent for the solubilization of cellular membranes.[26] The detergent-containing samples were subjected to centrifugation to pellet the membrane fragments, and the solubilized proteins were purified using Ni-NTA resin. The resulting elution fractions were then quantified using the MicroBCA assay and analyzed by SDS-PAGE (Figure 2-9). As described in Figure 2-9A, each archaeal OTase was successfully purified from the *E. coli* membrane in good yield, indicating a 2-8-fold

increase in total purified protein per liter of *E. coli* culture compared to PglB. However, the final yields of the archaeal OTases seemed to be lower than what was predicted from the Western blot analysis; this discrepancy may be due in part to incomplete solubilization of the archaeal OTases from the cell membrane by Triton X-100. Nonetheless, it is clear that all four archaeal OTases are expressed at a considerably higher level in *E. coli* compared to PglB.

A

OTase	Quantity of Purified OTase (μg/L *E. coli* culture)
M. voltae	549
M. maripaludis	312
A. fulgidus	193
H. marismortui	416
C. jejuni (PglB)	77

B

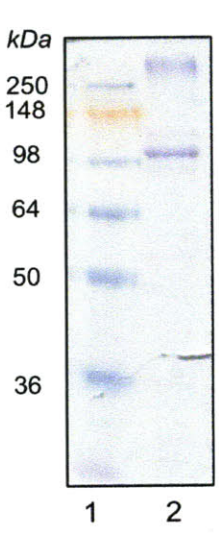
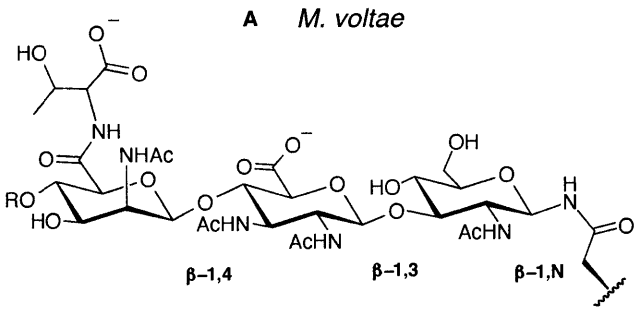


Figure 2-9: (A) Summary of the overall yield of purified archaeal OTases compared with PglB; (B) SDS-PAGE analysis of a purified sample of the *M. voltae* OTase. (1) MW standard; (2) *M. voltae* OTase.

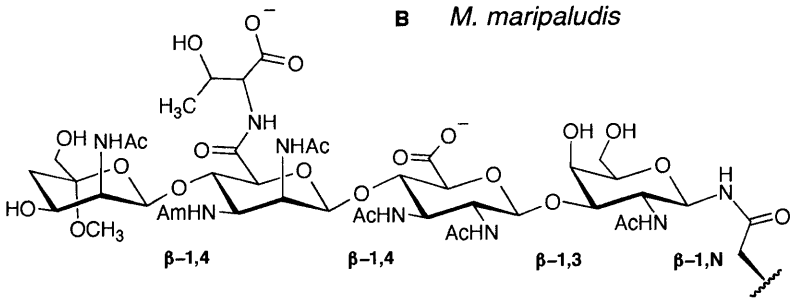
Testing Archaeal OTases for Function: Synthesis of Preliminary OTase Substrates

Based on the results described in the previous sections of this chapter, four archaeal OTases have been identified that exhibit a substantial increase in protein expression compared to PglB. In order to utilize these proteins for future in depth study, the function of these proteins must be demonstrated; thus, the next step of this work required the synthesis of potential OTase substrates. To date, only the structures of the *M. voltae* and *M. maripaludis* *N*-linked glycans have been elucidated (Figure 2-10), and nothing is known of the glycoproteins produced by *A.*

fulgidus or *H. marismortui*.[27,28] The *N*-linked glycan of *M. voltae*, depicted in Figure 2-10A, is a trisaccharide in which the linking sugar to asparagine is GlcNAc. The next two carbohydrates of this glycan are 2,3-dideoxy-2,3-diacetamido-glucuronic acid (GlcNAc(3NAc)A) and an *N*-acetylmannosaminuronic acid residue (ManNAc(6Thr)A), in which a threonine is linked to the C6" carboxylate. Neither of these two sugars has been previously identified in *N*-linked glycan structures. In *Methanococcus maripaludis*, the flagella are modified with a linear tetrasaccharide that is β-linked to the asparagine through a GalNAc residue (Figure 2-10B).[27] GlcNAc(3NAc)A is ligated to the proximal GalNAc sugar, followed by a highly modified mannosaminuronic (ManNAcA) residue that contains an acetamidino group at the C3" position and a threonine at the C6" carbon, similar to the *M. voltae* glycan. In addition, the terminal sugar of this oligosaccharide is the unique 2-acetamido-2,4-dideoxy-5-*O*-methyl-hexos-ulo-1,5-pyranose, which represents the first reported example of an aldulose in an *N*-linked glycan structure.[27] As the building blocks of these *N*-linked glycans are not commercially available nor predicted to be synthesized in a facile manner, we sought to first attempt functional studies using a simplified glycosyl donor and peptide acceptor; these efforts are outlined below.



ManNAc(6Thr)A-β-1,4-GlcNAc(3NAc)A-β-1,3-GlcNAc-β-1,N-Asn



2NAc(4deoxy)hexulose-β-1,4-ManNAc(3Am6Thr)A-β-1,4-GlcNAc(3NAc)A-β-1,3-GalNAc-β-1,N-Asn

Figure 2-10: *N*-linked glycans identified in (A) *M. voltae* and (B) *M. maripaludis*. The unique, highly-modified carbohydrates that comprise these glycans are highlighted in blue.

a. Synthesis of M. voltae Flagellar Peptides

As a first step toward probing the function the archaeal OTases, a panel of acceptor peptides was designed and synthesized based upon the glycoproteins identified in *M. voltae* (Figure 2-11). Analysis of these protein sequences indicates that unlike PglB, the *M. voltae* OTase does not appear to require an extended glycosylation sequence beyond the N-X-S/T motif found in eukaryotes. A series of six peptides was designed based on the S-layer and flagellar protein sequences, all of which contain three residues on either side of the N-X-S/T glycosylation sequence in order to optimize binding contacts with the archaeal OTases. In addition, these peptides are acetylated at the N-terminus and include a C-terminal *p*-

nitrophenylalanine residue (Nph) to aid in quantification. The peptides were synthesized on PAL-PEG-PS resin using standard Fmoc-solid phase peptide techniques, purified by RP-HPLC, and verified by mass spectrometry. The sequences of these peptides are shown in blue in Figure 2-11.

Fla A

ATGVGTLIVF IAMVLVAAVA ASVLI**NTS**GF LQQKASSTGT ESTEQVSTGL KMFQTSGKLN EPIIDRLTIY
VTPSPGSKPV DLKNTKLLMN RWTFQPPPVS YSSTYFENNN KQIFDVTGSK AWNNGAILPE YNFGVIVIQD
DDGSCTAESP VIGKGDMAVI TI**NCT**NLDLA PRTRLNGYLQ SEIGFKTQFT YILPNAYDKT EDVVILQ

Fla B1

ASGIGTLIVF IAMVLVAAVA ASVLI**NTS**GF LQQKASTTGK ESTEQVASGL QVTGVTGV**NS** **T**DSE**NIT**HIV
AYITPKAGSS AIDLSQAKLF VTYNGVSAVL KA**NES**AIDGT SGTPDVLSAT LLS**NTS**ATEY QVVSLQDFDG
SVAKNQVINK GDLVAIRIST GQVFSSTKGI PTRAHVSGKL QPEFGAPGII EFTTPATFTT KVVELQ

Fla B2

ASGIGTLIVF IAMVLVAAVA ASVLI**NTS**GF LQQKASTTGK ESTEQVASGL QISQVMGMH**N** **NS**NI**NKT**AIY
ISPNAGSSAI DLSQAVIMLS DGSNKRVYKY **NES**SYKDLTN GGDIFDNANV EWIKATATKF GIVVIQDADE
SCTAANPVIN KGDLVAITL**N** **TT**SFSTTPRT SITGTVQPEF GAPGIISFTT PATYL**NDS**KV VQLQ

Fla B3

AVGIGTLIIF IALVLVAAVA ASVIINTAGK LQHKAAVVGQ ETTQQVASGL QVVKITGHSV DQYNLDKIAI
LVSPNIGDEI DLATTVVTFS TDDRKMSLLY DSS**NAS**NGGK VRLSTA**NGT**S DIFKYDDVYA IGAWPFEDPT
YGQSDQDPNE KKFGIVVLQD MDNSVSGEHP TVNYGDKVLL AINIGNIVGE NIGNTITIQG EVVPEFGAPG
IIDFTTPPVY

S-layer Protein

VEKIGDVEGF KVIDNGEPTA DIVVGSTAAA ADVVSAANVA AKVGSMMFKE GEAASGSAKL TVKASAESDD
ANLKSLLT**NG** **T**NDFTELDAG KEAFVVAAAD SDYSDALI**NA** **T**TGFANIADN VLYDQAKLAA AVSLGDLSTL
SVVKDIDPSD WYADKNKAAD VATKDYYDQD GDAVEMLMAT VASNDDGKSL TVDEDGVLYA SIAYDDDNED
FQRATQVLKE GNRLPFLGEE YALVKLDTDD DIVYLGKEVF DGVLKEGDTY NIGDGYELKV VAILKSGDEY
KISLQMKDG KVVAEKFDKV SATSALKMIY TPGNIGIVVN EAWENVGQDY GYGSTLITKD VIALELGEEY
IPDWEVVTIE KDTTTDNTKD SKMTLSDDKI TKDNTYGIGL QYVGDEEDNF KSGKAIKIAK YAELELDDED
KEDTKLNLFF SMDETKEATL AAGQKVTVLN SDITLSEVMA DAKAPVAFKA PLAVLDTEVS LDAANKKLIL
VGGPVANALT KELADAGKIE MTVESPATLA VVAGAANGND VLVVAGGDRA ATAEAANALI EML

Figure 2-11: Sequences of *M. voltae* flagella and S-layer glycoproteins. The boldface residues indicate sites where *N*-linked glycosylation was observed; the sequences highlighted in blue represent peptides synthesized for OTase studies. Figure modified from Voisin et al.[28]

b. Synthesis of Dolichyldiphosphate [^{3}H]Chitobiose (Dol-PP-[^{3}H]GlcNAc$_2$)

In addition to peptide substrates, a suitable glycosyl donor was required to probe the archaeal OTases for function. Due to the complexity of the *M. voltae* and *M. maripaludis* *N*-linked glycan structures (Figure 2-10), a simpler disaccharide substrate was sought for use in the preliminary examination of enzyme function. Previous work in the Imperiali laboratory has shown that both the eukaryotic OT and the bacterial PglB are able to accept disaccharide substrates in the place of the full length oligosaccharides, and that the transfer of these truncated substrates to acceptor peptides is carried out with nearly complete turnover by the respective OTases.[15,29] Although the exact structure of the *M. voltae* polyisoprenyl-linked glycosyl donor is unknown, genetic studies in which the *M. voltae algH* gene was able to complement a conditionally lethal mutant of the analogous phosphoglycosyltransferase gene in *S. cerevisiae*, *alg7,* suggest that the polyisoprenyl carrier in *M. voltae* is dolichyldiphosphate.[30] Therefore, the synthesis of Dol-PP-[^{3}H]GlcNAc$_2$ was carried out; this compound bears a tritium label on the terminal GlcNAc residue and has been routinely used in the Imperiali laboratory for the study of the yeast OT.

Previous studies have determined that the most efficient method for the synthesis of Dol-PP-[^{3}H]GlcNAc$_2$ involves treatment of Dol-PP-GlcNAc with freshly prepared porcine liver microsomes and UDP-[^{3}H]GlcNAc (Figure 2-12), as expression and in vitro reconstitution of the Alg13/14 complex in *S. cerevisiae* has not yet been achieved. To this end, a liver from a freshly slaughtered pig was obtained and rapidly processed to prepare microsomes, which were immediately used to carry out the synthesis of Dol-PP-[^{3}H]GlcNAc$_2$. The crude reaction mixture was purified by silica gel chromatography to yield pure Dol-PP-[^{3}H]GlcNAc$_2$, which was verified by testing against the yeast OT complex.

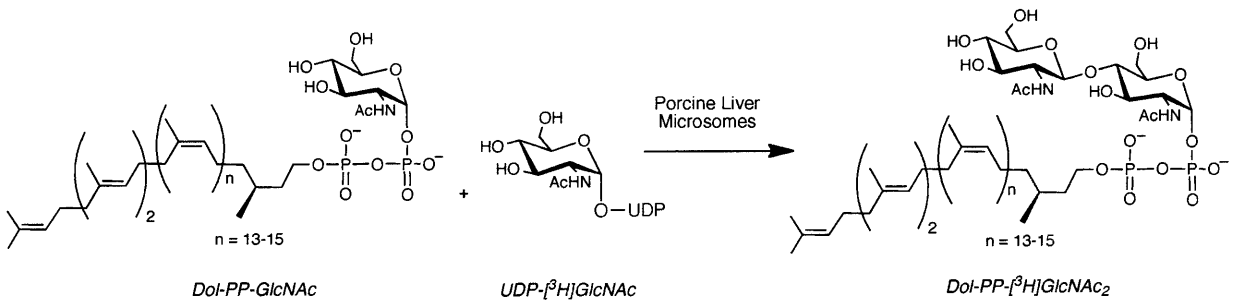


Figure 2-12: Chemoenzymatic synthesis of Dol-PP-[^{3}H]GlcNAc$_2$ using freshly prepared porcine liver microsomes.

Testing the Archaeal OTases for Function

With both the radiolabeled Dol-PP-GlcNAc$_2$ substrate and peptide acceptor sequence in hand, the archaeal OTases from *M. voltae, M. maripaludis, A. fulgidus*, and *H. marismortui* were screened for function. These proteins were overexpressed in *E. coli* as described previously and crude membrane fractions were prepared. Blank membrane fractions in which both BL21 Gold and RIL cells were transformed with an empty pET24a(+) vector were assayed as well; these samples served as controls for background activity that may be inherent to the *E. coli* cell membrane fraction.

A standard assay was used to test these archaeal OTases for activity, based upon that originally developed for the yeast OT (Figure 2-13).[15,31] It involves the incubation of the OTase with the radiolabeled disaccharide donor and other reaction components, including detergent and $MnCl_2$. The assay is initiated with the addition of the peptide substrate, and aliquots are removed at various time points and quenched in an aqueous/organic mixture, which serves to partition the glycosylated peptide into the aqueous phase and the polyisoprenyl-starting material in the organic phase. Glycosylation of the peptide substrate by the OT homolog would result in

a net transfer of radioactive counts from the organic phase into the aqueous phase as measured by scintillation counting.

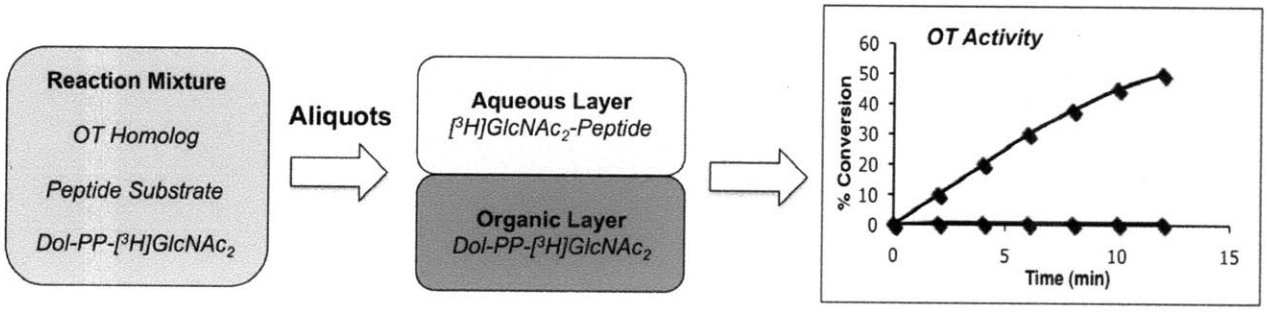


Figure 2-13: Standard assay for OTase activity. The archaeal OTases are incubated with the peptide substrate, the radiolabeled glycan donor, and other reaction components and aliquots are removed over time and subjected to phase extraction. Glycosylation results in the transfer of radioactive counts from the organic to the aqueous phase.

Unfortunately, none of the archaeal OTases tested showed appreciable activity above the background (Figure 2-14). While the radioactivity present in the reaction aliquots increased over time, this increase was later attributed to inherent glycosidase activity present in *E. coli* cell membrane, which is supported by the observation of considerable activity in the empty pET vector control sample. Further analysis of the reaction mixtures by HPLC and mass spectrometry did not show any evidence of glycopeptide formation. These assays were repeated numerous times, in which addition of various divalent metals, salts, detergents and reducing agents were included in the reaction. The temperature and pH of the reactions were also altered; however, disaccharide transfer to protein was never observed. The Und-PP-Bac[^{3}H]GalNAc substrate of PglB was also utilized as a possible glycosyl donor in these reactions, but no activity was observed in these reactions either.

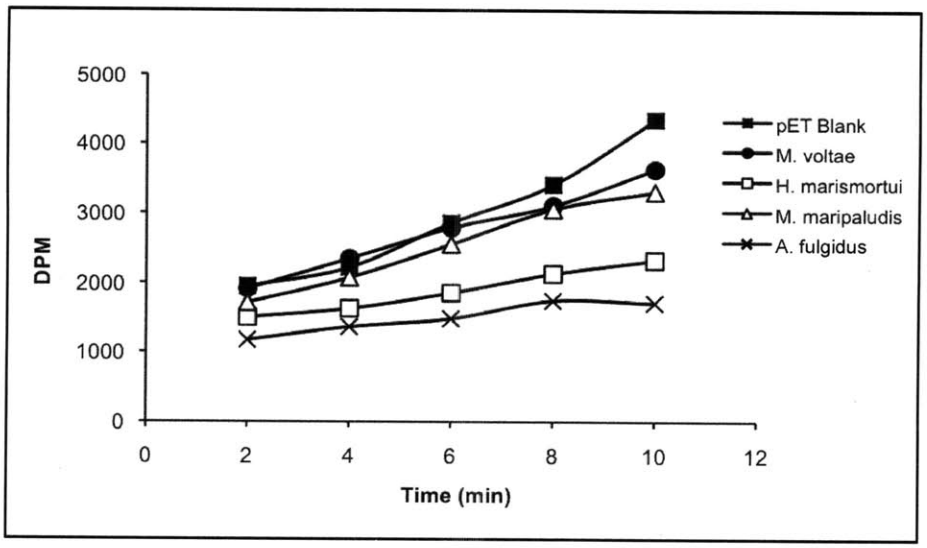


Figure 2-14: Assay data from a representative archaeal OTase activity assay, in which Dol-PP-$(GlcNAc)_2$ was used as the glycosyl donor and Ac-YKYNESSYKNph-NH_2 was introduced as the acceptor peptide.

As an alternative to disaccharide Dol-PP-$GlcNAc_2$, the monosaccharide Dol-PP-GlcNAc was also tested as a possible archaeal OTase substrate. It was hypothesized that the eukaryotic disaccharide donor, in which the two carbohydrates are connected through a β-1,4 linkage, may not able to efficiently bind to the *M. voltae* or *M. maripaludis* OTase as the native *N*-linked glycans in these species feature a β-1,3 linkage between the first two sugars. In order to test this hypothesis, the standard OTase assay was performed using Dol-PP-GlcNAc as the glycosyl donor (Figure 2-15). However, use of this substrate is complicated by the lack of a radiolabel in the transferred sugar. This difficulty was circumvented through analysis of the OTase reaction by HPLC and mass spectrometry to identify the presence of glycopeptide.

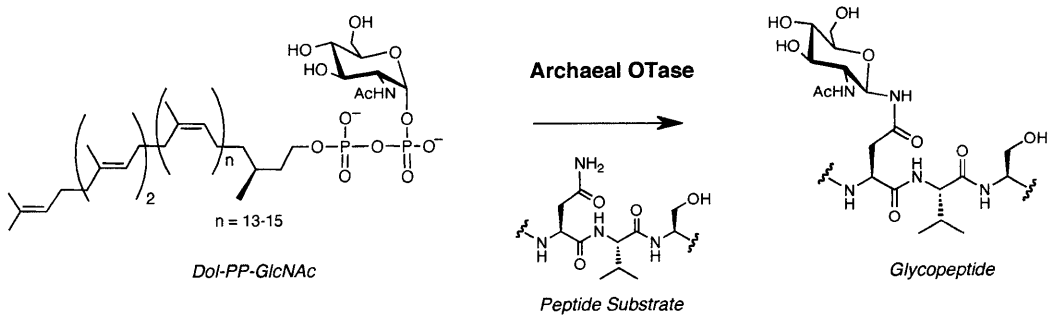


Figure 2-15: Archaeal OTase assay, in which the formation of the glycopeptide product is detected through the use of HPLC and mass spectrometry.

In order to test the monosaccharide as an archaeal OTase substrate, the standard assay was employed with slight modification. The reaction components were assembled and allowed to incubate for 8 hrs, after which the entire reaction was quenched and subjected to aqueous/organic phase separation as previously described. Rather than analyze the aqueous layer by scintillation counting, the entire mixture was injected onto a C_{18} RP-HPLC column, with the goal of separating the glycopeptide product from the peptide starting material. As the addition of a single sugar was not expected to influence the retention time of the peptide a great deal, an extremely shallow HPLC gradient was employed, in which the elution solvent was introduced over the course of 90 mins. However, analysis of each OTase reaction yielded only a single peak on the HPLC corresponding to peptide (Figure 2-16). Mass spectrometry revealed only the presence of the peptide starting material, suggesting that glycosylation had not occurred. A shorter peptide, Ac-NLTNph-NH_2, was also utilized in the OTase reactions in an effort to aid in the HPLC separation of the potential product from the starting material. Although this short peptide was expected to make fewer binding contacts with the OTase, thus resulting in lower glycosylation efficiency, it was proposed that the addition of a single sugar to

this peptide would result in a more dramatic shift in retention time. However, as in the case of the longer decapeptides, formation of the desired glycopeptides was not observed.

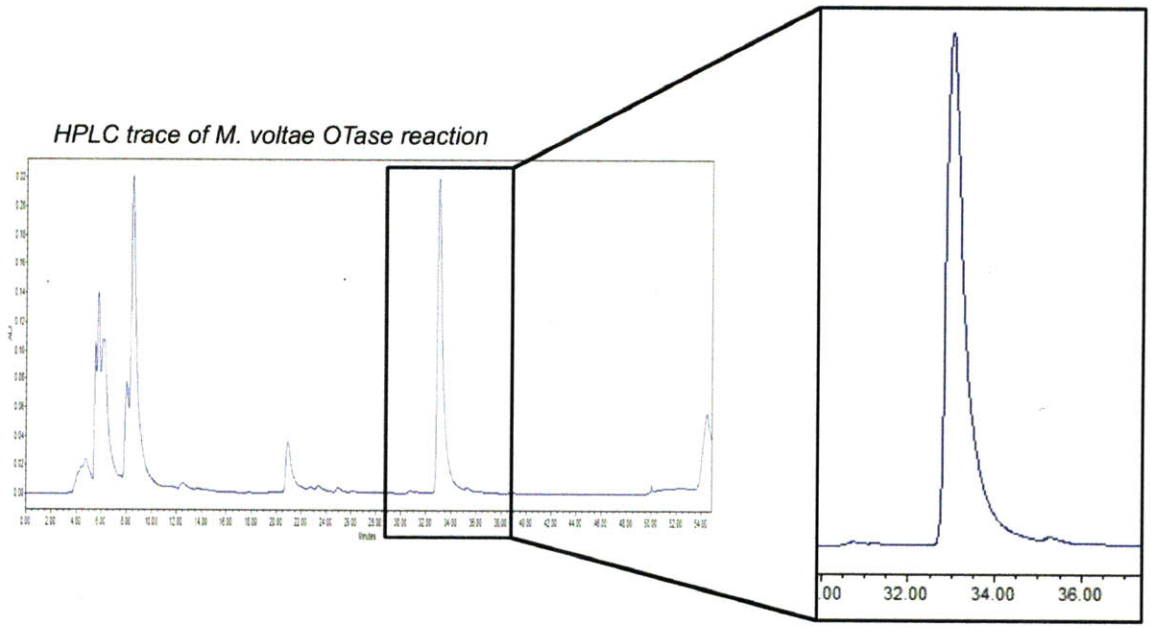


Figure 2-16: HPLC trace of the *M. voltae* OTase assay. The highlighted peak corresponds to the peptide starting material; no evidence of the glycopeptide product was observed by mass spectrometry.

Next Steps: Synthetic Strategy for Generation of a Suitable Archaeal OTase Substrate

As previously described, the archaeal OTases identified in this chapter were unable to catalyze glycosylation of the acceptor peptides using either Dol-PP-GlcNAc or Dol-PP-$GlcNAc_2$ as glycosyl donors. Therefore, in order to probe these OTases for function, it was proposed that a polyisoprenyl-linked glycan bearing greater similarity to the native *M. voltae* and *M. maripaludis* *N*-linked glycans must be synthesized. A retrosynthetic strategy for the synthesis of two dolichyl-linked glycans, Dol-PP-GlcNAc-GlcNAc(3NAc)A and Dol-PP-

GalNAc-GlcNAc(3NAc)A, is outlined in Figure 2-17 and involves two key transformations. First, the novel nucleotide sugar UDP-GlcNAc(3NAc)A is prepared by either chemical or chemoenzymatic means, followed by the coupling of this sugar to Dol-PP-GlcNAc or Dol-PP-GalNAc using a suitable glycosyltransferase.

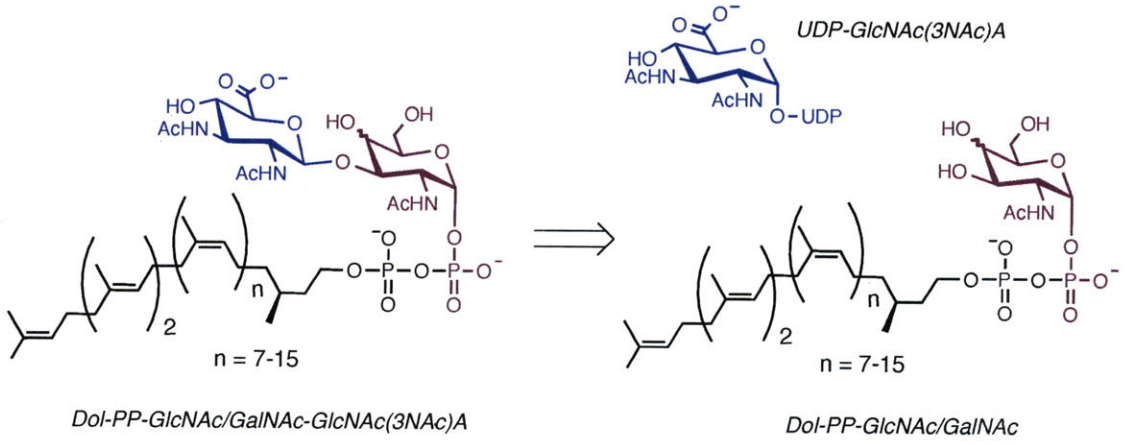


Figure 2-17: Retrosynthetic strategy for the generation of the desired dolichyl-linked disaccharides, Dol-PP-GlcNAc-GlcNAc(3NAc)A and Dol-PP-GalNAc-GlcNAc(3NAc)A.

In the first step of this strategy, a robust method for the synthesis of UDP-GlcNAc(3NAc)A is required. Rather than obtain this compound through chemical synthesis, a laborious process that has since been shown to entail over fifteen steps culminating in a low final yield,[32] we sought out a set of biosynthetic enzymes from the Wbp pathway in the Gram-negative opportunistic pathogen *Pseudomonas aeruginosa* PAO1. These enzymes have been implicated in the assembly of the unique carbohydrate ManNAc(3NAc)A, which is a critical component of the B-band O-antigen that comprises the lipopolysaccharide (LPS) of the organism (Figure 2-18). Upon generation of UDP-GlcNAc(3NAc)A, efforts to identify a suitable glycosyltransferases to ligate this sugar to either Dol-PP-GlcNAc or Dol-PP-GalNAc is required. To this end, a panel of glycosyltransferases from both *M. voltae* and *M. maripaludis*

has been identified, and progress towards the synthesis of the final dolichyl-linked disaccharides is summarized in Chapter 5.

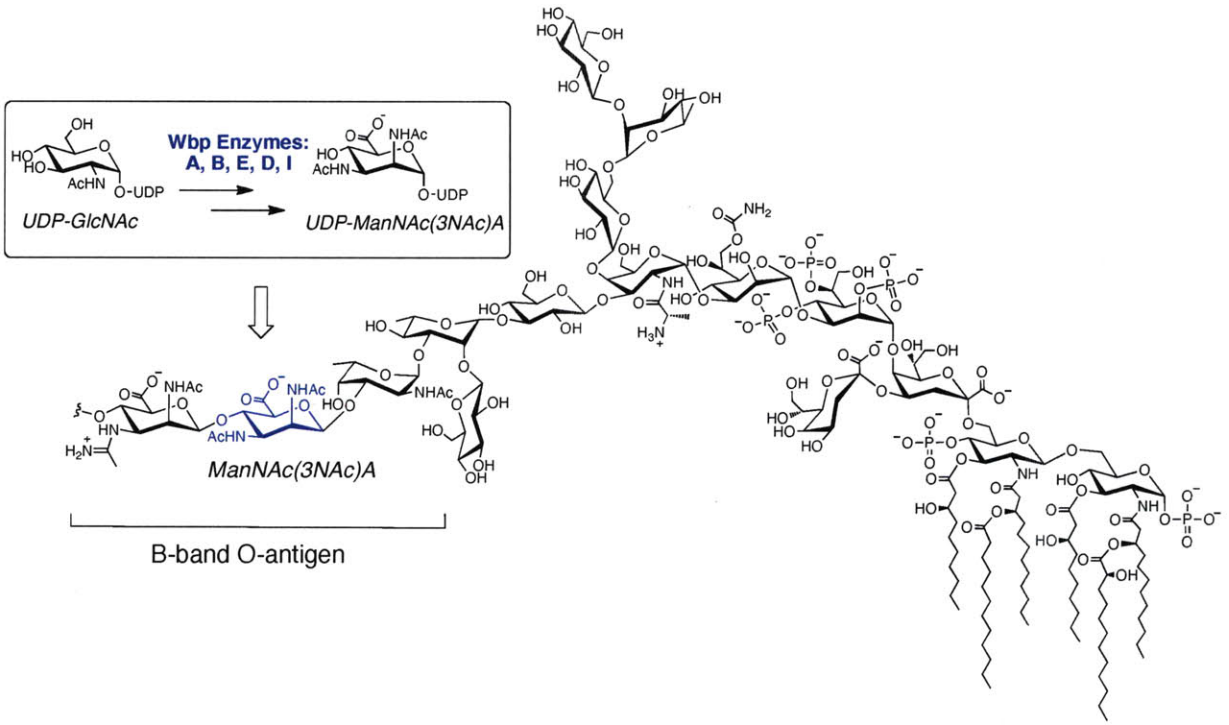


Figure 2-18: The lipopolysaccharide (LPS) of *Pseudomonas aeruginosa* PAO1 (O5), with the ManNAc(3NAc)A moiety in the B-band O-antigen highlighted in blue. The enzymes WbpA, WbpB, WbpE, WbpD, and WbpI have been implicated in the biosynthesis of UDP-ManNAc(3NAc)A, and the desired UDP-GlcNAc(3NAc)A is proposed to be an intermediate in this pathway. The characterization of these Wbp enzymes is outlined in Chapter 3.

Conclusions

In this chapter, efforts towards the identification and characterization of an archaeal OTase are described. An initial search of completed archaeal genome sequences yielded a total of 25 archaeal Stt3 homologs, from which 5 candidates were selected for further analysis. After cloning these genes into suitable vectors, protein expression levels were compared in various *E. coli* cell lines by Western blot analysis. These studies indicated that the OTases from the

methanogen *Methanococcus voltae* and the extreme halophile *Haloarcula marismortui* exhibited the highest overall protein expression yields, representing a nearly 100-fold improvement over PglB. Further work demonstrated that these proteins could be purified from the cell membrane; however, efforts to observe function of these proteins using various glycosyl donors were unsuccessful. It is hypothesized that the lack of activity demonstrated by these archaeal OTases is due to the use of substrates that differ in carbohydrate linkage and composition from the native archaeal glycosyl donors; thus, future work will entail the biosynthesis of a dolichyl-linked substrate containing the unique GlcNAc(3NAc)A. Efforts towards this goal are summarized in Chapters 3 and 5 of this thesis. In order to begin to unravel the enzymatic mechanism of the oligosaccharide transfer reaction, a long-standing interest of the Imperiali laboratory, large quantities of a highly pure and active OTase homolog are required, and the work described in this chapter represents the first steps towards this goal.

Acknowledgements

I am grateful to Professor W. Whitman (University of Georgia, Athens, GA) and Professor K.F. Jarrell (Queen's University, Kingston, Ontario) for providing genomic DNA from *M. jannaschii*, *M. voltae,* and *M. maripaludis*, and UROP students Yunji Wu and Minah Shahbaz for early help with molecular biology. I would also like to thank Meredith Hartley and Professor Barbara Imperiali for assistance with preparation of crude porcine liver microsomes, Michelle Chang for providing the Ac-NLTNph-NH_2 peptide, and Dr. Cliff Stains and Dr. Andrew Krueger for analysis of peptides by MALDI.

Experimental Methods

General Information

Unless otherwise noted, all solvents and reagents were obtained commercially and used without further purification. Genomic DNA was purchased from the American Culture Type Collection (Manassas, VA) or generously provided by Professors W. Whitman (University of Georgia, Athens) and K.F. Jarrell (Queen's University). Oligonucleotides were purchased from Eurofin MWG Operon (Huntsville, AL) or Sigma Life Sciences (St. Louis, MO). Restriction endonucleases were obtained from New England Biolabs (Ipswich, MA), and competent cell lines are from Stratagene (Agilent Technologies, Santa Clara, CA). Sequencing of all bacterial plasmids was conducted by the MIT CCR Biopolymers Laboratory (Cambridge, MA). Fmoc-protected amino acids and coupling agents were purchased from Novabiochem (Gibbstown, NJ) and GenScript (Piscataway, NJ). Solvents were obtained from VWR (West Chester, PA) or Sigma Aldrich (St. Louis, MO). Fmoc-PAL-PEG-PS resin was acquired from Applied Biosystems (Carlsbad, CA). HPLC was carried out on a Waters Prep LC 4000 or Waters Delta 600 systems, each equipped with a Waters 2487 dual wavelength detector. For analytical HPLC, a C_{18} YMC ODS-A 5 μm (4.6 x 250 mm) column was used, and for preparatory HPLC, a C_{18} YMC-Pack ODS-A 5 μm (250 x 20 mm) column was employed. ESI-MS was conducted on a Mariner instrument (Applied Biosystems).

Cloning of Archaeal stt3 Homologs

The *stt3* homolog genes were amplified by the polymerase chain reaction from genomic DNA from the following organisms and strains: *Methanocaldococcus jannaschii* DSM,

Methanococcus voltae, *Methanococcus maripaludis* S2, *Archaeoglobus fulgidus* (ATCC 490203), and *Haloarcula marismortui* (ATCC 43049) using the Pfu Turbo polymerase (Stratagene) and the oligonucleotides listed in Appendix (Table 1). The resulting PCR products containing both *Bam*HI and *Xho*I restriction sites (or *Bam*HI and *Hind*III for *H. marismortui*) were digested and cloned into the same sites of the pET24a(+) vector (Novagen) via standard molecular biology techniques. The final gene products encoded proteins with an N-terminal T7-tag and a C-terminal His_6-tag.

Overexpression of Archaeal OTases

The pET24a(+) plasmids containing the desired archaeal *stt3* homologs were transformed into the following *E. coli* competent cell lines: BL21(DE3) Gold, BL21-CodonPlus(DE3) RIL, BL21-CodonPlus(DE3) RP, and BL21(DE3) pLysS (Stratagene), using the appropriate antibiotics for selection. For overexpression, 1 L of Luria-Bertani media supplemented with the corresponding antibiotics was inoculated at 37 °C while shaking until an optical density (600 nm) of 0.6-0.8 was achieved. The cultures were then cooled to 16 °C and protein expression was induced by the addition of IPTG (1 mM). After 16 hrs, the cells were harvested by centrifugation (5000 x *g*) and the resultant cell pellets were stored at -80 °C until needed.

Preparation of Crude E. coli Membrane Fractions

All steps were performed at 4 °C. A frozen cell pellet of *E. coli* expressing an archaeal Stt3 homolog was resuspended in 50 mM HEPES, pH 8.0 / 1 mM EDTA buffer (30 mL) supplemented with Protease Inhibitor Cocktail III (Calbiochem, 30 μL). The cells were lysed by

sonication, and the mixture was centrifuged at 6000 x *g* for 30 mins to remove the cellular debris. The supernatant was gently decanted and subjected to a second round of centrifugation (140,000 x *g*) for 1 hr to collect the membrane fraction. The pellet was washed with 50 mM HEPES, pH 8.0 / 1 mM EDTA buffer, centrifuged (140,000 x *g*) for 1 hr, then homogenized in 50 mM HEPES, pH 8.0 / 100 mM NaCl buffer (3 mL). The crude cell membrane fraction was divided into 100 μL aliquots and stored at -80 °C for future use.

Purification of Archaeal OTases

All steps were performed at 4 °C. A 500 μL aliquot of a cell envelope fraction containing an archaeal Stt3 homolog was removed from storage and thawed on ice, after which Triton X-100 was added to a final concentration of 1%. The mixture was incubated for 12 hrs with gentle rocking, after which it was subjected to centrifugation (140,000 x *g*) to pellet the membranes. The supernatant was removed and incubated with Ni-NTA agarose resin (Qiagen) for 2 hrs with gentle rocking, then subsequently poured into a fritted PolyPrep column (BioRad) to collect the resin. The resin was washed (50 mM HEPES, pH 8.0/300 mM NaCl/25 mM imidazole) and the protein eluted (50 mM HEPES, pH 8.0/300 mM NaCl/250 mM imidazole). Fractions containing the desired product were identified by SDS-PAGE and Western blot, using an Anti-T7 antibody for detection. Protein concentration was determined by either the Micro BCA kit (Pierce) or UV absorbance using the appropriate extinction coefficient.

Synthesis of M. voltae Flagellar Peptides

All peptides were synthesized using standard solid phase Fmoc-based protocols on Fmoc-PAL-PEG-PS resin (0.19 mmol/g). The resin was swelled first in CH_2Cl_2 (5 mins) then DMF (5 mins) prior to synthesis. Deprotection of the terminal Fmoc group was achieved by exposure of the resin to an excess of 20% 4-methylpiperidine in DMF (3 x 5 mins). Each amino acid was coupled to the resin by incubation with the amino acid (4 equiv), PyBOP (4 equiv), and DIPEA (8 equiv) for 45 mins. Between each deprotection and coupling step, the resin was washed with DMF (5 x 1 min) then CH_2Cl_2 (5 x 1 min). The TNBS test was used to check the resin for coupling efficiency, in which the formation of red color indicates the presence of a free amine.[33] All peptides were acetylated at the N-terminus by incubation with acetic anhydride (10 equiv) and pyridine (10 equiv) in DMF for 30 mins. The side chain protecting groups were removed and the peptides cleaved from the resin by exposure to a TFA:CH_2Cl_2:H_2O:TIS mixture (90:5:2.5:2.5) for 3 hrs while shaking vigorously, followed by filtration to remove the beads and solvent evaporation under a stream of nitrogen. The resulting pellet was triturated with cold Et_2O, and the peptides were purified by preparative C_{18} RP-HPLC using a 45 minute gradient of 5-95% mobile phase B, where the mobile phases are H_2O / 0.1% TFA (A) and CH_3CN / 0.1% TFA (B). Purified peptides were characterized by analytical RP-HPLC for purity and either ESI-MS or MALDI-TOF MS for identity. Peptides were stored as solutions of DMSO and quantified by UV absorbance using the ε_{280} of Nph (12,500 $M^{-1}cm^{-1}$).

Preparation of Crude Porcine Liver Microsomes

All steps were performed at 4 °C. A liver from a freshly slaughtered pig (E.M. Blood Company, Groton, MA) was obtained and sliced into four 100 g sections, mincing each section into small

chunks. Using a blender, 200 g of liver was combined with 200 mL homogenizing buffer (50 mM TrisOAc, pH 7.2 / 250 mM sucrose / 2.5 mM $MgCl_2$ / 3 mM DTT / 0.1 mM AEBSF) and mixed until smooth. A slow speed spin was then performed (4,500 x g) to pellet organ debris, after which the supernatant was decanted into fresh centrifuge tubes and spun again. The supernatant was then transferred into clean tubes again and centrifuged (140,000 x g) to pellet the membranes. Each pellet was resuspended in 300 μL homogenizing buffer, adding glycerol to a final concentration of 30%. The mixture was divided into 250 μL aliquots and stored at -80 °C for future use.

Chemoenzymatic Synthesis of Dolichyldiphosphate [^{3}H]Chitobiose (Dol-PP-[^{3}H]-GlcNAc$_2$)

Dolichyldiphosphate *N*-acetylglucosamine (1 mg) was dissolved in 3:2 $CHCl_3$:MeOH (166.66 μL) to a concentration of 6 mg/mL, and 60 μL of this solution was added to each of 10 clean glass scintillation vials and dried under a stream of N_2. A 1:10 dilution of UDP-*N*-[^{3}H]acetylglucosamine (10 uCi) was prepared in 70% EtOH, and 10 μL of this solution was added to each vial containing the dolichyl-sugar and dried. DMSO (50 μL) and freshly prepared conversion buffer (800 μL, 50 mM TrisOAc, pH 7.0 / 250 mM sucrose / 5 mM $MgCl_2$ / 3 mM DTT / 1% NP-40) were subsequently added to each vial and vortexed. Freshly thawed porcine liver microsomes (150 μL) were added to this mixture and gently shaken at room temperature for 30 mins. The reactions were quenched by the addition 3:2:1 $CHCl_3$:MeOH:4 mM $MgCl_2$ (6 mL), and each vial was centrifuged to achieve separation of the organic and aqueous phases. The aqueous layer was removed, and the organic layer was twice extracted with theoretical upper layer (TUP, 3 mL). The aqueous phases were then recorded and quantified with

scintillation counting to determine overall reaction turnover. The organic layers were combined and dried, and purified with silica gel (72:21:3 $CHCl_3$:MeOH:2.5 mM $MgCl_2$). Each fraction was analyzed by scintillation counting, and upon the appearance of radioactivity, the column was washed 60:25:4 $CHCl_3$:MeOH:2.5 mM $MgCl_2$. The fractions containing the desired product were combined and quantified, and stored in 50,000 DPM aliquots at -20 °C until further use.

Oligosaccharyl Transferase Assays

To a tube of dried Dol-PP-[^{3}H]$GlcNAc_2$, DMSO (10 μL), 2X assay buffer (100 μL, 100 mM HEPES, pH 7.5 / 280 mM sucrose / 2.4% v/v Triton X-100), 1 M $MnCl_2$ (2 μL), H_2O (53 μL), and freshly thawed archaeal OTase crude membranes were combined. The assay was initiated by the addition of 10 μL of a 2 mM stock of the peptide of interest in DMSO. Aliquots of the reaction mixture (25 μL) were removed every 2 mins and quenched by addition of 3:2:1 $CHCl_3$:MeOH: 4mM $MgCl_2$ (1.2 mL). The aqueous layer was then removed and the organic layer extracted with theoretical upper phase (TUP) with salt (2 x 600 μL). The aqueous layers were combined, mixed with 5 mL of scintillation fluid (Ecolite, MP Biomedicals) and subjected to scintillation counting. In the case of assays involving Dol-PP-GlcNAc, the aqueous layers were injected onto an analytical C_{18} RP-HPLC column and purified extended version of the gradient described above, in which the elution solvent is introduced from 0-50% over the course of 90 mins. The resulting peaks were collected, dried down, and analyzed by MALDI.

References

1. Varki, A.; Cummings, R.D.; Freeze, H.H.; Stanley, P.; Bertozzi, C.R.; Hart, G.W.; Etzler, M.E. *Essentials of Glycobiology*; 2nd Ed.; Cold Spring Harbor Laboratory Press: New York, 2008.
2. Weerapana, E.; Imperiali, B. Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **2006**, *16*, 91R-101R.
3. Jarrell, K.F.; Jones, G.M.; Nair, D.B. Biosynthesis and role of N-linked glycosylation in cell surface structures of archaea with a focus on flagella and S-layers. *Int. J. Microbiol.* **2010**, *2010*, 470138.
4. Nothaft, H.; Szymanski, C.M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **2010**, *8*, 765-778.
5. Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **1983**, *209*, 331-336.
6. Lehle, L.; Strahl, S.; Tanner, W. Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 6802-6818.
7. Dempski Robert, E., Jr.; Imperiali, B. Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr Opin Chem Biol* **2002**, *6*, 844-850.
8. Kelleher, D.J.; Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **2006**, *16*, 47R-62R.
9. Pathak, R.; Hendrickson, T.L.; Imperiali, B. Sulfhydryl modification of the yeast Wbp1p inhibits oligosaccharyl transferase activity. *Biochemistry* **1995**, *34*, 4179-4185.
10. Schulz, B.L.; Stirnimann, C.U.; Grimshaw, J.P.A.; Brozzo, M.S.; Fristsch, F.; Mohorko, E.; Capitani, G.; Glockshuber, R.; Grütter, M.G.; Aebi, M. Oxidoreductase activity of oligotransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11061-11066.
11. Yan, Q.; Lennarz, W.J. Studies on the function of oligosaccharyl transferase subunits. *J. Biol. Chem.* **2002**, *277*, 47692-47700.
12. Li, H.; Chavan, M.; Schindelin, H.; Lennarz, W.J.; Li, H. Structure of the oligosaccharyl transferase complex at 12 Å resolution. *Structure* **2008**, *16*, 432-440.
13. Szymanski, C.M.; Yao, R.; Ewing, C.P.; Trust, T.J.; Guerry, P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **1999**, *32*, 1022-1030.
14. Wacker, M.; Linton, D.; Hitchen, P.G.; Nita-Lazar, M.; Haslam, S.M.; North, S.J.; Panico, M.; Morris, H.R.; Dell, A.; Wren, B.W.; Aebi, M. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **2002**, *298*, 1790-1793.
15. Glover, K.J.; Weerapana, E.; Numao, S.; Imperiali, B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **2005**, *12*, 1311-1315.
16. Nita-Lazar, M.; Wacker, M.; Schegg, B.; Amber, S.; Aebi, M. The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **2005**, *15*, 361-367.

17. Chen, M.M.; Glover, K.J.; Imperiali, B. From peptide to protein: comparative analysis of the substrate specificity of *N*-linked glycosylation in *C. jejuni*. *Biochemistry* **2007**, *46*, 5579-5585.
18. Feldman, M.F.; Wacker, M.; Hernandez, M.; Hitchen, P.G.; Marolda, C.L.; Kowarik, M.; Morris, H.R.; Dell, A.; Valvano, M.A.; Aebi, M. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3016-3021.
19. Glover, K.J.; Larkin, A.; Imperiali, B. Unpublished work, 2006.
20. Vieille, C.; Zeikus, G.J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1-43.
21. Mescher, M.F.; Strominger, J.L. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *J. Biol. Chem.* **1976**, *251*, 2005-2014.
22. Calo, D.; Kaminski, L.; Eichler, J. Protein glycosylation in archaea: sweet and extreme. *Glycobiology* **2010**, *20*, 1065-1076.
23. Lechner, J.; Wieland, F.; Sumper, M. Biosynthesis of sulfated saccharides N-glycosidically linked to the protein via glucose. Purification and identification of sulfated dolichyl monophosphoryl tetrasaccharides from halobacteria. *J. Biol. Chem.* **1985**, *260*, 860-866.
24. Varki, A.; Cummings, R.D.; Freeze, H.H.; Stanley, P.; Bertozzi, C.R.; Hart, G.W.; Etzler, M.E. *Essentials of Glycobiology*; 2nd Ed.; Cold Spring Harbor Laboratory Press: New York, 2009.
25. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes *J. Mol. Biol.* **2001**, *305*, 567-580.
26. Newby, Z.E.R.; O'Connell, J.D.; Gruswitz, F.; Hays, F.A.; Harries, W.E.C.; Harwood, I.M.; Ho, J.D.; Lee, J.K.; Savage, D.F.; Miercke, L.J.W.; Stroud, R.M. A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nat. Protocols* **2009**, *4*, 619-637.
27. Kelly, J.; Logan, S.M.; Jarrell, K.F.; VanDyke, D.J.; Vinogradov, E.V. A novel *N*-linked flagellar glycan from *Methanococcus maripaludis*. *Carb. Res.* **2009**, *344*, 648-653.
28. Voisin, S.; Houliston, R.S.; Kelly, J.; Brisson, J.-R.; Watson, D.; Bardy, S.L.; Jarrell, K.F.; Logan, S.M. Identification and characterization of the unique *N*-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *J. Biol. Chem.* **2005**, *280*, 16586-16593.
29. Imperiali, B.; Hendrickson, T.L. Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.* **1995**, *3*, 1565-1578.
30. Shams-Eldin, H.; Chaban, B.; Niehus, S.; Schwarz, R.T.; Jarrell, K.F. Identification of the archaeal *alg7* gene homolog (encoding *N*-acetylglucosamine-1-phosphate transferase) of the *N*-linked glycosylation system by cross-domain complementation in *Saccharomyces cerevisiae*. *J. Bacteriol.* **2008**, *190*, 2217-2220.
31. Sharma, C.B.; Lehle, L.; Tanner, W. N-Glycosylation of yeast proteins. Characterization of the solubilized oligosaccharyl transferase. *Eur. J. Biochem.* **1981**, *116*, 101-108.
32. Rejzek, M.; Kannathasan, V.S.; Wing, C.; Preston, A.; Westman, E.L.; Lam, J.S.; Naismith, J.H.; Maskell, D.J.; Field, R.A. Chemical synthesis of UDP-Glc-2,3-diNAcA,

a key intermediate in cell surface polysaccharide biosynthesis in the human respiratory pathogens *B. pertussis* and *P. aeruginosa*. *Org. Biomol. Chem.* **2009**, *7*, 1203-1210.

33. Hancock, W.S.; Battersby, J.E. A new micro-test for the detection of incomplete coupling reactions in solid-phase peptide synthesis using 2,4,6-trinitrobenzene-sulphonic acid. *Anal. Biochem.* **1976**, *71*, 260-264.

Chapter 3 Biosynthesis of UDP-GlcNAc(3NAc)A Using Enzymes in the Wbp Pathway of *Pseudomonas aeruginosa* PAO1

A significant portion of the work described in this chapter has been published in the following:

Larkin, A. and Imperiali, B. Biosynthesis of UDP-GlcNAc(3NAc)A by WbpB, WbpE, and WbpD: Enzymes in the Wbp Pathway Responsible for O-Antigen Assembly in *Pseudomonas aeruginosa* PAO1. *Biochemistry* **2009**, *48*, 5446-5455.

Introduction

The Gram-negative pathogen *Pseudomonas aeruginosa* is a versatile organism responsible for infection in immunocomprised individuals.[1] It is a leading source of hospital-acquired pneumonia and bacteremia, causes severe inflammation and pulmonary failure in cystic fibrosis patients, and has emerged as a serious public health threat.[2-5] Effective treatment of *P. aeruginosa* infection has proved challenging due to the strong inherent resistance of the organism to traditional antibiotics and the increasing emergence of multidrug resistant strains.[6-8] While several vaccines for *P. aeruginosa* have been described, none thus far have achieved clinical success.[9]

One of the major factors affecting the virulence of *P. aeruginosa* is the composition of the unique carbohydrates that make up the lipopolysaccharide (LPS).[4,10] Localized in the exterior leaflet of the outer membrane of the organism, the *P. aeruginosa* LPS is composed of three distinct regions: lipid A, which anchors the structure to the membrane, a core oligosaccharide, and the O-antigen, a strand of monosaccharides that is further classified as either A-band or B-band (Figure 3-1).[11-13] Unlike the A-band O-antigen, which is a homopolymer of α-1,3-linked D-rhamnose,[14] the B-band O-antigen is structurally complex and can vary between strains; this diversity serves as the basis of serological classification of particular strains of the organism.[15,16] In addition, the B-band O-antigen has been shown to play a critical role in host colonization and provides resistance to both serum sensitivity and phagocytosis.[12,17,18] In *P. aeruginosa* PAO1 (serotype O5), the B-band O-antigen is composed of repeating units of a trisaccharide containing 2-acetamido-3-acetamidino-2,3-dideoxy-β-D-mannuronic acid (ManNAc(3Nam)A), 2,3-diacetamido-2,3-dideoxy-β-D-mannuronic acid (ManNAc(3NAc)A), and *N*-acetyl-α-D-fucosamine (Fuc2NAc).[19] Interestingly, it has been

shown that while all three of these carbohydrates are derived from a common precursor, UDP-GlcNAc, the biosynthesis of each involves a distinct set of genes.[20]

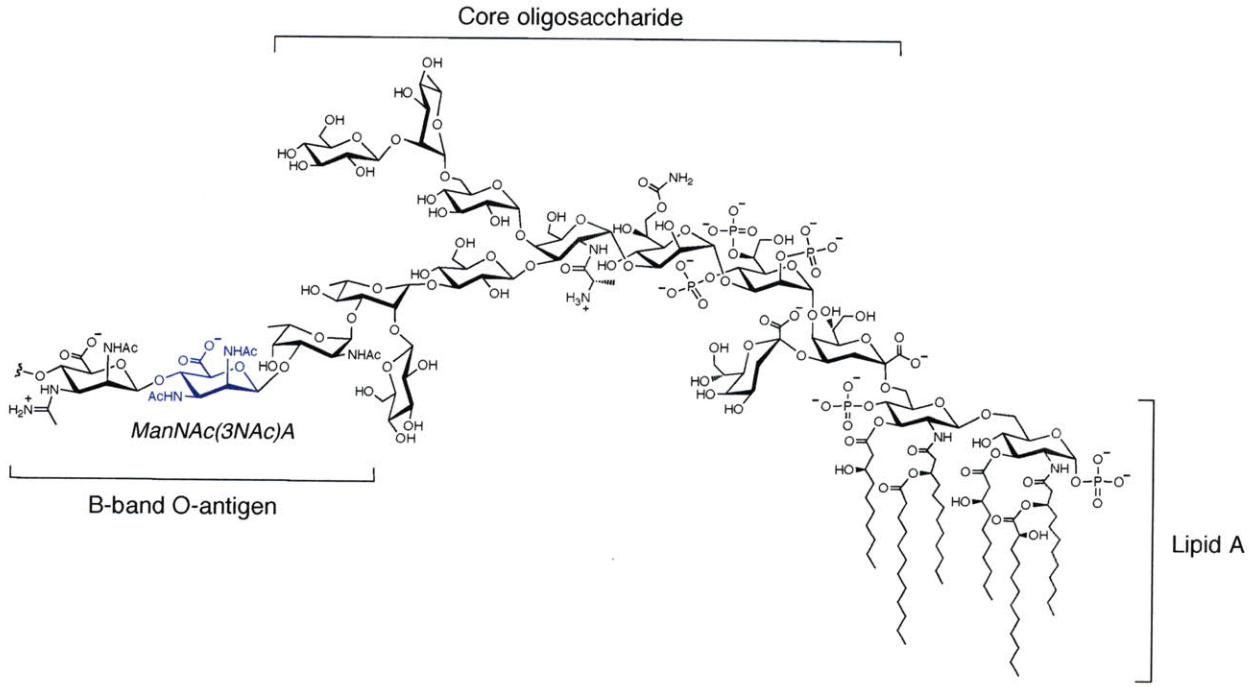


Figure 3-1: General structure of the lipopolysaccharide of *P. aeruginosa* PAO1 (serotype O5), depicted with one unit of the B-band form of the O-antigen. ManNAc(3NAc)A is highlighted in blue.

The assembly of the B-band O-antigen trisaccharide takes place on the cytoplasmic surface of the inner membrane of *P. aeruginosa*, and involves the sequential transfer of each carbohydrate unit onto an undecaprenyl-phosphate carrier (Und-P, Figure 3-2).[12,21] A series of genetic experiments suggest that the first sugar, Fuc2NAc, is added by the glycosyltransferase WbpL, followed by the addition of ManNAc(3NAc)A and ManNAc(3NAm)A by WbpH and WbpJ.[22] After synthesis of the trisaccharide is complete, it is processed in a similar manner to that described for LPS assembly in *Escherichia coli*.[23] Following this model, the O-antigen trisaccharide unit is first flipped into the periplasm by the Wzx translocase, an integral

membrane protein conserved in all Gram-negative bacteria.[24] Individual trisaccharide units are joined together by the Wzy polymerase, which works in concert with the chain length regulator Wzz to control the extent of polymerization. In *P. aeruginosa* PAO1, B-band O-antigen chain length may vary from 12 to over 100 units in size.[12,25] The means by which Wzy and Wzz coordinate to control O-antigen chain length is an area currently under investigation; preliminary evidence suggests that a coiled coil extended from the surface of Wzz interacts with the growing O-antigen produced by Wzy, resulting in the regulated chain length found in each serotype of *P. aeruginosa*.[26-28] The completed O-antigen is then transferred to the nascent LPS by the WaaL ligase, followed by the shuttling of the entire LPS from the periplasm to the outer membrane through a complex comprised of LptD and LptE.[29]

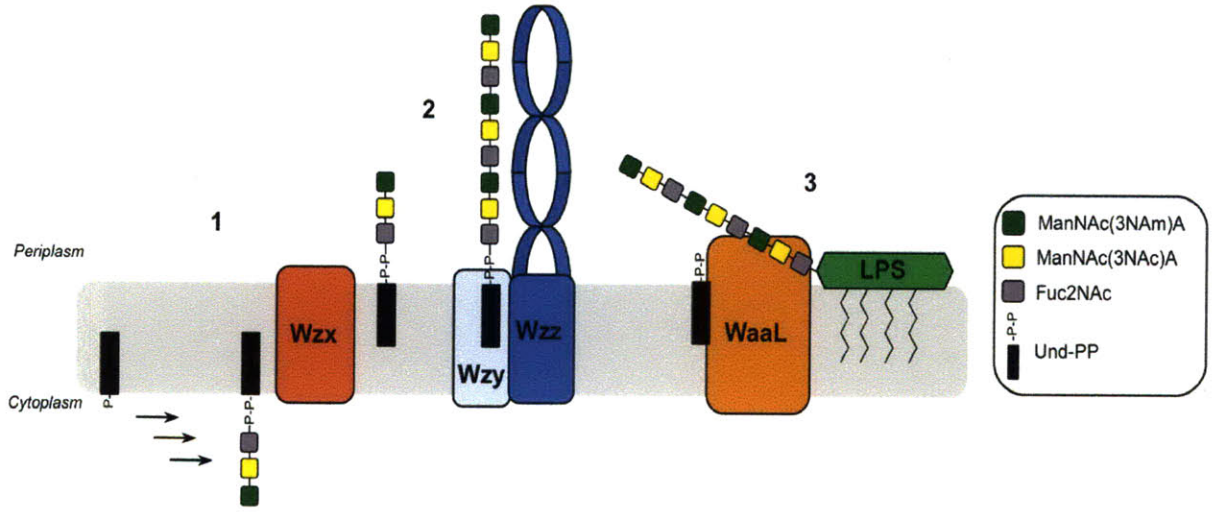


Figure 3-2: General scheme of B-band O-antigen biosynthesis in *Pseudomonas aeruginosa* PAO1. 1) The O-antigen trisaccharide is synthesized on the cytoplasmic face of the inner membrane onto an Und-P carrier, then flipped to the periplasm by the Wzx translocase; 2) the O-antigen chain is lengthened by the Wzy polymerase, controlled by the chain regulator Wzz; 3) the completed O-antigen is tethered to the nascent LPS by the WaaL ligase.

Previous work involving a combination of genetic and biochemical analyses have indicated that the UDP-activated form of the second sugar in the B-band O-antigen, UDP-

ManNAc(3NAc)A, is the product of genes in the Wbp pathway (Figure 3-3). The enzymes encoded by these genes are thought to convert UDP-GlcNAc to UDP-ManNAc(3NAc)A in a stepwise fashion, followed by the transfer of the ManNAc(3NAc)A moiety onto an undecaprenyl carrier by the putative glycosyltransferase WbpH. The biosynthetic pathway begins with WbpA, previously shown to catalyze the C6"-oxidation of UDP-GlcNAc to give the corresponding UDP-*N*-acetyl-D-glucosaminuronic acid (UDP-GlcNAcA).[30] It is then hypothesized that the C3"-dehydrogenase WbpB, aminotransferase WbpE, and acetyltransferase WbpD sequentially convert UDP-GlcNAcA into UDP-2,3-diacetamido-2,3-dideoxy-D-glucuronic acid (UDP-GlcNAc(3NAc)A), which is the UDP-activated form of a key carbohydrate found in both the *Methanococcus voltae* and *Methanococcus maripaludis N*-linked glycans.[31,32] Finally, the C2"-epimerase WbpI modifies UDP-GlcNAc(3NAc)A to give the final UDP-ManNAc(3NAc)A. Genetic mutants of *P. aeruginosa* in which *wbpA*, *wbpB*, *wbpE*, *wbpD* and *wbpI* have been deleted lack the B-band O-antigen, thus highlighting their critical importance to pathogenicity.

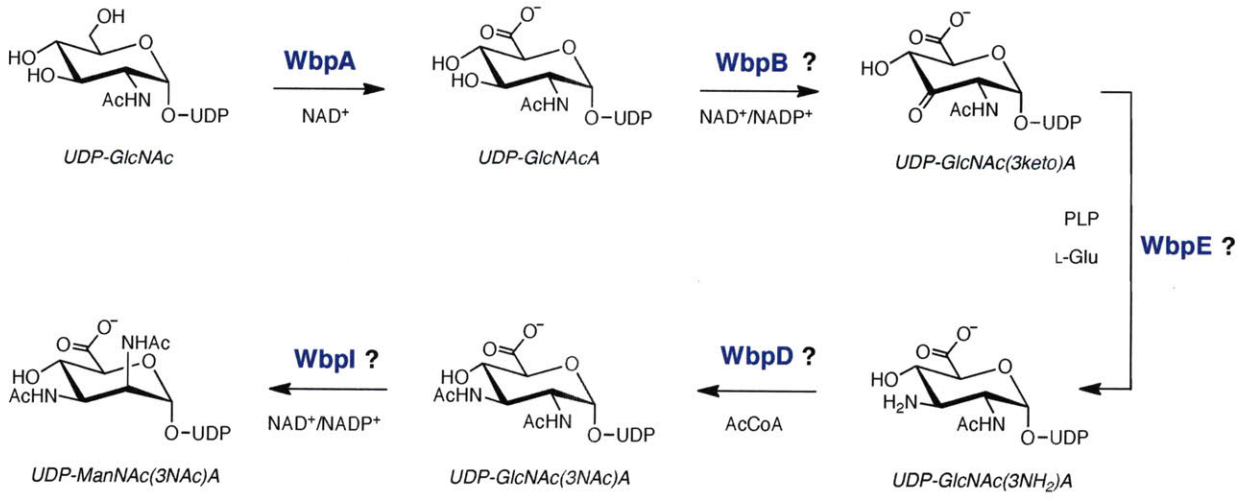


Figure 3-3: Biosynthetic pathway of UDP-ManNAc(3NAc)A in *Pseudomonas aeruginosa*, as defined prior to this study.

As outlined in Chapter 2, our efforts to define the *N*-linked glycosylation pathway in the organisms *M. voltae* and *M. maripaludis* require a robust means for generating UDP-GlcNAc(3NAc)A. Rather than obtain this compound through chemical synthesis, a laborious process that has since been shown to entail over fifteen steps culminating in a low final yield,[33] we sought to characterize three enzymes from *P. aeruginosa* PAO1, WbpB, WbpE, and WbpD, implicated in the biosynthesis of the desired nucleotide sugar. In this chapter, we present the overexpression, purification, and biochemical characterization of these three enzymes and demonstrate that they are responsible for the biosynthesis of UDP-GlcNAc(3NAc)A. We also identify a novel NAD^+ recycling mechanism which requires the coupling of WbpB and WbpE, the dehydrogenase and aminotransferase, to give the WbpE product UDP-GlcNAc($3NH_2$)A. Notably, WbpB, WbpE, and WbpD can be used in conjunction with WbpA and WbpI to convert UDP-GlcNAc to UDP-ManNAc(3NAc)A in a one-pot reaction. This work completes the characterization of the UDP-ManNAc(3NAc)A biosynthetic pathway in *P. aeruginosa* and provides, for the first time, a straightforward synthetic route to multimilligram quantities of the rare nucleotide sugar UDP-GlcNAc(3NAc)A, a critical intermediate for future studies of LPS assembly and the oligosaccharyl transferase in methanogenic archaea.

Results and Discussion

Overexpression and Purification of WbpA, WbpB, WbpE, and WbpD

The *wbpA*, *wbpB*, *wbpE*, and *wbpD* genes were amplified from *P. aeruginosa* PAO1 genomic DNA and inserted into suitable plasmids for protein expression in *E. coli* using standard molecular biology techniques. The resulting proteins were overexpressed in high yield;

15 mg (WbpA), 20 mg (WbpB), 110 mg (WbpE), and 15 mg (WbpD) of purified protein was routinely obtained from 1 L cell culture. The expected molecular weights of WbpA, WbpB, WbpE, and WbpD, each containing an N-terminal T7 and C-terminal His_6-tag, corresponded to the observed molecular weights based on SDS-PAGE and Western blot analysis, using antibodies directed against the T7 and His_6-tags (Figure 3-4). The enzymes retained activity when stored at -20 °C (WbpA, WbpB, and WbpE) or 4 °C (WbpD) for a minimum of three months.

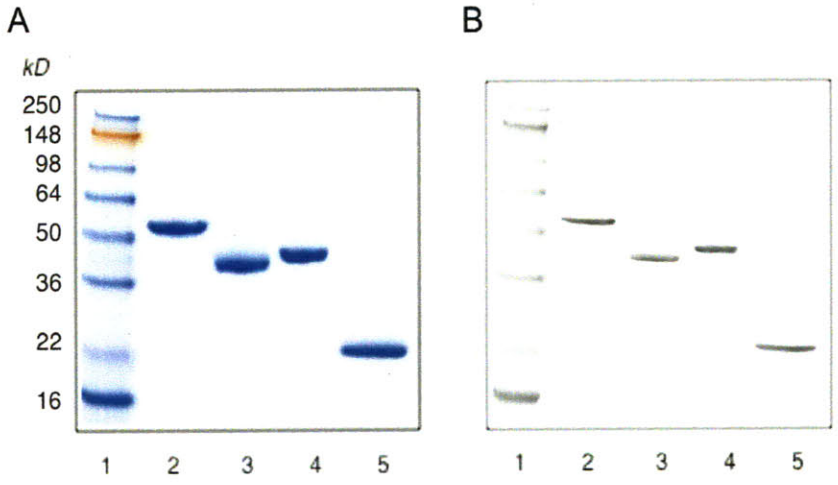


Figure 3-4: (A) 10-20% gradient SDS-PAGE and (B) Anti-T7 Western blot. (1) MW standard; (2) WbpA; (3) WbpB; (4) WbpE; (5) WbpD.

Synthesis of UDP-N-Acetylglucosaminuronic acid (UDP-GlcNAcA)

The first step in generating the desired UDP-GlcNAc(3NAc)A involved the synthesis of the glucuronic acid, UDP-GlcNAcA. In order to probe the function of the remaining enzymes in the Wbp pathway, a direct method for generating large quantities of this nucleotide sugar was required. An initial attempt to synthesize this compound entailed the use of WbpA, the first enzyme in the biosynthetic pathway, which had been previously characterized by Miller et al.[30]

WbpA, a 50.3 kD protein containing one N-terminal transmembrane domain, is homologous to the UDP-GlcNAc 6-dehydrogenases from *Achromobacter georgiopolitanum* and *Micrococcus luteus* and has been shown to catalyze the oxidation of UDP-GlcNAc to UDP-GlcNAcA, utilizing 2 molecules of NAD^+ to convert the C6" hydroxyl group to a carboxylic acid (Figure 3-5).[34,35] Kinetic analysis of WbpA revealed that the reaction takes place at an optimal pH of 8.5 and requires the presence of monovalent cations, in particular NH_4^+ or K^+.[30] These kosmotropic ions are believed to stabilize the protein and serve to enhance the reaction rate. A crystal structure of the *E. coli* fructose-1,6-bisphosphate aldolase, which also requires these cations for activity, shows a common binding site for both ions;[36] it is hypothesized that WbpA may contain a homologous cation binding site, which would explain the requirement of NH_4^+ or K^+ for enzyme activity.

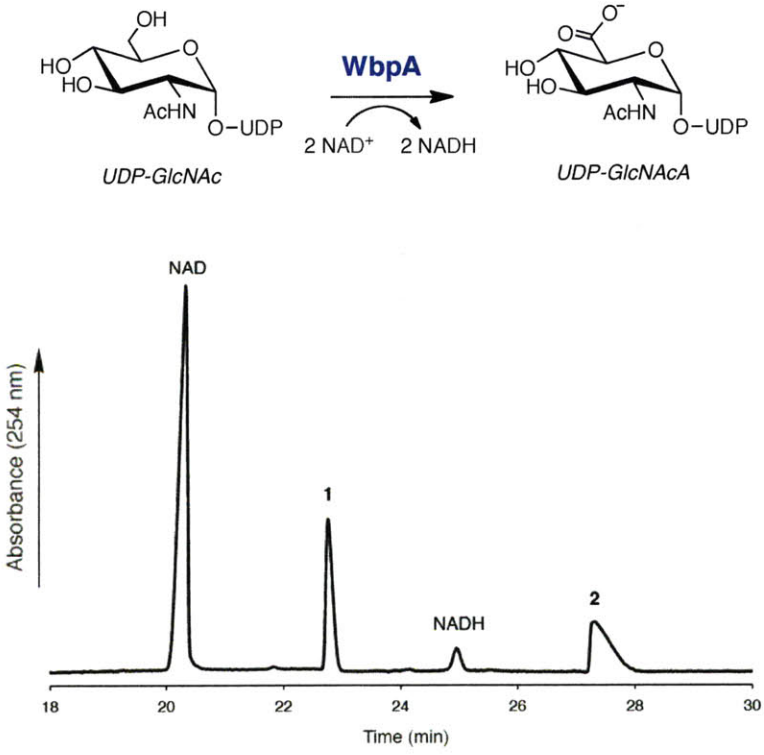


Figure 3-5: Capillary electrophoresis chromatogram showing conversion of UDP-GlcNAc (1) to UDP-GlcNAcA (2) by WbpA.

To this end, WbpA was incubated with NAD^+ and UDP-GlcNAc, and reaction progress was monitored by capillary electrophoresis (CE, Figure 3-5). Although formation of the UDP-GlcNAcA product was observed, the overall substrate conversion was modest (31%). Further optimization of conditions, including modifying reaction temperature, additive concentrations, and quantity of enzyme, provided a marginal increase in overall reaction yield but was still not sufficient to afford full conversion of substrate to product. It was hypothesized that maximal substrate turnover did not take place due to inhibition of WbpA by the product, UDP-GlcNAcA. Rather than struggle with efforts to overcome this obstacle, we sought an alternative means to obtain our desired compound.

Recently, Field and coworkers reported a direct oxidation of UDP-GlcNAc to UDP-GlcNAcA in the presence of platinum and oxygen without the use of protecting groups.[37] While this selective oxidation has been previously demonstrated,[38,39] this work represents the first time the methodology has been applied to UDP-GlcNAc and thus provides a useful alternative to the low-yielding chemoenzymatic route. The reaction involves streaming oxygen gas into a stirring mixture of UDP-GlcNAc and the activated platinum catalyst in H_2O at reflux over the course of three days (Figure 3-6). After removal of the platinum catalyst by filtration, the reaction mixture is subjected to anion exchange chromatography to separate the product from the remaining starting material, followed by further purification by C18 reversed-phase HPLC to desalt the material and remove small impurities. Optimization of the reaction conditions, including adjustment of the oxygen flow rate and vigorous stirring of the reaction mixture, has afforded a reaction yield of nearly 75% on a routine basis, therefore providing a robust means of generating pure UDP-GlcNAcA for our studies.

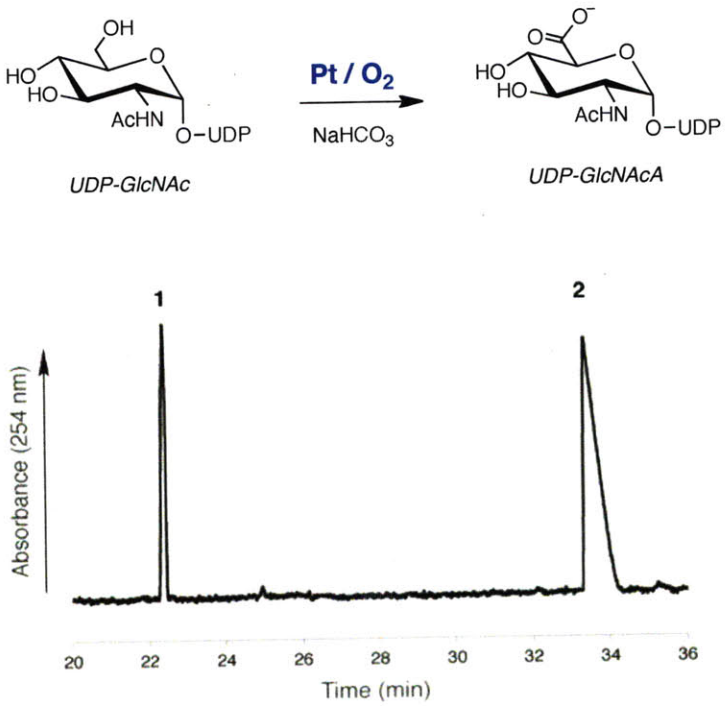


Figure 3-6: CE chromatogram depicting the Pt/O_2-catalyzed oxidation of UDP-GlcNAc (1) to UDP-GlcNAc (2).

Functional Characterization of the Coupled WbpB and WbpE Reaction

With a reliable source of UDP-GlcNAcA in hand, efforts were then made to characterize the function of the next enzymes in the Wbp pathway, WbpB and WbpE. WbpB, a previously uncharacterized 35.7 kD protein, was predicted to be a dehydrogenase responsible for oxidation of the C3" hydroxyl group to afford the ketone product. Primary sequence analysis revealed that WbpB is a putative member of the GFO/IDH/MocA family of oxidoreductases, containing a characteristic $NAD^+/NADP^+$-binding Rossmann fold.[40] However, a search for homologs of WbpB in the NCBI database only resulted in a few matches, including uncharacterized proteins from various strains of *P. aeruginosa* and *Bordetella pertussis*. The next enzyme in the pathway, WbpE, is a 38.7 kD protein found to be a member of the broad class of Fold Type I aspartate

aminotransferases. This class of proteins utilizes the cofactor pyridoxal 5'-phosphate (PLP) and an amino acid donor to convert a carbonyl group into a primary amine. A search for WbpE homologs in the NCBI database resulted in over twenty related nucleotide sugar aminotransferases, including the well-studied proteins DesV (*Streptomyces venezuelae*) and PseC (*Helictobacter pylori*).[41,42]

Initial attempts to characterize the function of WbpB in the presence of UDP-GlcNAcA and NAD^+ or $NADP^+$ were unsuccessful despite an exhaustive screen of reaction conditions. Efforts included the addition of various monovalent and divalent cations (Na^+, K^+, NH_4^+, Ca^{2+}, Mg^{2+}, Mn^{2+}), salts (NaH_2PO_4, K_2HPO_4, $(NH_4)_2SO_4$, KCl), cofactors (NAD^+, $NADP^+$, FAD), detergents (Triton X-100) and reducing agents (DTT), as well as varying the reaction pH (6.0-8.5) and temperature (25-37 °C). Despite these attempts, no formation of the UDP-GlcNAc(3keto)A product was observed by CE. In order to account for the unlikely event that UDP-GlcNAc(3keto)A was indeed formed but was co-eluted with the starting material UDP-GlcNAcA on the CE, benzylamine was used as a ketone-labeling agent and added to the reaction mixture after two hours. If the ketone product was generated, it would react with benzylamine at a basic pH to form the imine adduct and thus be predicted to elute from the CE at a different retention time. However, incubation of the reaction mixture with benzylamine did not result in any evidence of WbpB catalysis.

As a last effort to test for activity of WbpB, UDP-GlcNAcA and NAD^+ were incubated with the next enzyme in the pathway, the aminotransferase WbpE, along with PLP and L-glutamate. Surprisingly, complete turnover of the UDP-GlcNAcA starting material and the formation of a new peak were observed by CE (Figure 3-7). The treatment of WbpE alone with UDP-GlcNAcA and obligate cofactors did not result in product formation, implying that both

enzymes were required for catalysis. The product of the reaction was later identified to be UDP-GlcNAc($3NH_2$)A by mass spectrometry and NMR analysis. In addition, the use of ^{15}N-L-glutamate as the amine donor resulted in a product with an increase in mass of 1 amu, indicating that the ^{15}N-label was successfully transferred to the molecule. Further investigation of the coupled WbpB/WbpE reaction showed that the enzymes catalyzed product formation at a wide range of pH (5.5-10.0) and temperature (4-65 °C). In addition, exogenous NAD^+ or PLP was not required for product formation, suggesting that both WbpB and WbpE were purified with their respective cofactors already bound. However, the addition of 200 μM NAD^+ and 100 μM PLP to the reaction mixture aided in achieving complete turnover of substrate, which implies that the heterologously expressed proteins were not saturated with cofactor due to the limiting intracellular levels of both NAD^+ and PLP in *E. coli*. Optimal reaction conditions yielding complete conversion of starting material to product was observed at pH 8.0 and 30 °C, and have led to the generation of over 20 mg of the desired UDP-GlcNAc($3NH_2$)A in a single reaction.

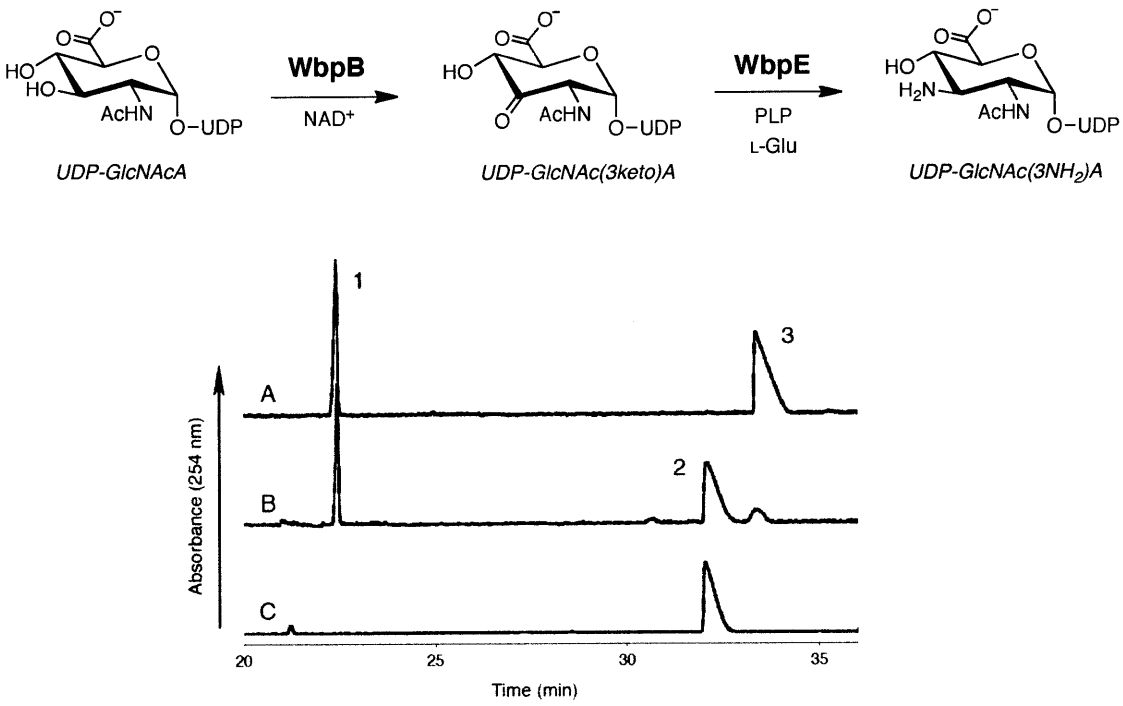


Figure 3-7: Capillary electrophoresis (CE) chromatogram representing (A) WbpB reaction alone, indicating no substrate conversion; (B) crude coupled WbpB/WbpE reaction; (C) pure UDP-GlcNAc(3NH_2)A. (1) NAD^+; (2) UDP-GlcNAc(3NH_2)A; (3) UDP-GlcNAcA.

The nucleotide sugar specificity of the coupled WbpB/WbpE reaction was explored by incubating the enzymes with UDP-GlcNAc, UDP-GalNAc, and UDP-D-glucuronic acid (UDP-GlcA) in the place of UDP-GlcNAcA. No turnover was observed in the presence of UDP-GlcNAc or UDP-GalNAc, and only minimal turnover was detected (11%) when UDP-GlcA was used as the nucleotide sugar substrate. These results confirm that WbpB prefers the glucopyranose configuration of the sugar as well as the presence of both the carboxylate at C6" and the acetylated amine at the C2" position. To analyze the amine donor specificity of WbpE, the WbpB/WbpE coupled reaction was screened with all 20 naturally occurring L-amino acids as well as D-glutamate, at a concentration of 25 mM. Other than L-glutamate, only the use of L-glutamine resulted in conversion from starting material to product (10%). This result strongly

suggests that WbpE is specific for L-glutamate as the amine donor, a result that has been confirmed for other homologous nucleotide sugar-modifying aminotransferases.[43,44]

Identification of the Nicotinamide Cofactor Bound to WbpB

In order to further characterize the coupled WbpB/WbpE reaction, we sought to determine whether WbpB utilizes either NAD^+ or $NADP^+$ to catalyze the oxidation of UDP-GlcNAcA. WbpB was precipitated by treatment with $HClO_4$ and the pellet washed several times to recover the bound cofactor. After neutralization and lyophilization, the extract was treated with either NAD^+-specific alcohol dehydrogenase or $NADP^+$-specific isocitrate dehydrogenase and the formation of NADH/NADPH monitored by UV absorbance at 340 nm.[45,46] As depicted in Figure 3-8, a sharp increase in absorbance was observed upon the addition of alcohol dehydrogenase, indicating the presence of NAD^+ in the extract. Alcohol dehydrogenase is specific for NAD^+, as suggested by the lack of absorbance increase when $NADP^+$ was introduced. In addition, no absorbance increase was seen after treatment of the extract with isocitrate dehydrogenase, which serves as further evidence that the bound cofactor to WbpB is NAD^+.

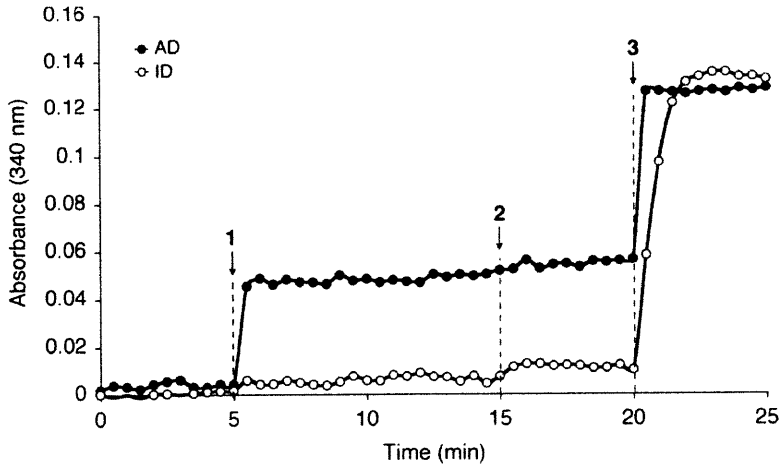


Figure 3-8: Identification of NAD^+ as the bound cofactor to WbpB by UV analysis. WbpB extract was treated with either NAD^+-dependent alcohol dehydrogenase (AD, *closed circles*) or $NADP^+$-dependent isocitrate dehydrogenase (ID, *open circles*) and analyzed for the formation of NADH/NADPH by absorbance at 340 nm. (1) Enzyme added to extract in cuvette; (2) $NADP^+$ (AD) or NAD^+ (ID) added to reaction to address specificity; (3) NAD^+ (AD) or $NADP^+$ (ID) added to reaction to check for activity.

Elucidation of a Novel Pathway for NAD^+ Recycling by WbpB

Perhaps the most striking finding obtained from the studies described above was that exogenous NAD^+ was not required for product formation by the coupled WbpB/WbpE reaction. In general, NAD^+-dependent nucleotide sugar dehydrogenases catalyze oxidation reactions through the abstraction of a hydride from the carbohydrate, followed by hydride transfer to NAD^+ to form NADH. However, in the case of the WbpB/WbpE coupled reaction, the formation of an equimolar amount of NADH was not observed in the reaction mixture, suggesting that WbpB efficiently recycles its NAD^+ cofactor. NAD^+ recycling by nucleotide sugar dehydrogenases has been widely observed; for example, in the case of the nucleotide sugar 4,6-dehydratases, such as Rff (*E. coli*), the oxidation of the C4" hydroxyl group results in NADH formation, followed by elimination of H_2O to generate an α, β-unsaturated ketone (Figure 3-9). Delivery of the hydride from NADH to the C6" carbon is coupled with protonation

at the C5" carbon to regenerate NAD^+ and prime the enzyme for another round of catalysis.[47,48] Unlike in this instance, the UDP-GlcNAcA substrate is not reduced over the course of the coupled WbpB/WbpE reaction, which suggests that the required oxidant is already present in reaction mixture. Therefore, we sought to identify the oxidant for NAD^+ recycling and proposed that it might in fact be α-ketoglutarate (α-KG), a product of the WbpE reaction.

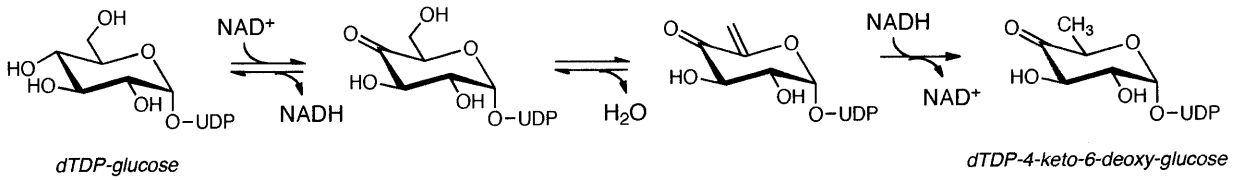


Figure 3-9: Reaction carried out by the dTDP-glucose 4,6-dehydratase from *E. coli*, Rff, in which NAD^+ is regenerated.[47]

In order to determine whether the required oxidant for NAD^+ regeneration was indeed α-KG, WbpB was incubated with UDP-GlcNAcA and 10 mM α-KG and the reaction monitored by CE. Remarkably, complete conversion of the UDP-GlcNAcA starting material to the UDP-GlcNAc(3keto)A product was observed in the presence of α-KG (Figure 3-10). No turnover was seen in the absence of either α-KG or WbpE, indicating the addition of α-KG to the WbpB was sufficient to promote NAD^+ recycling.

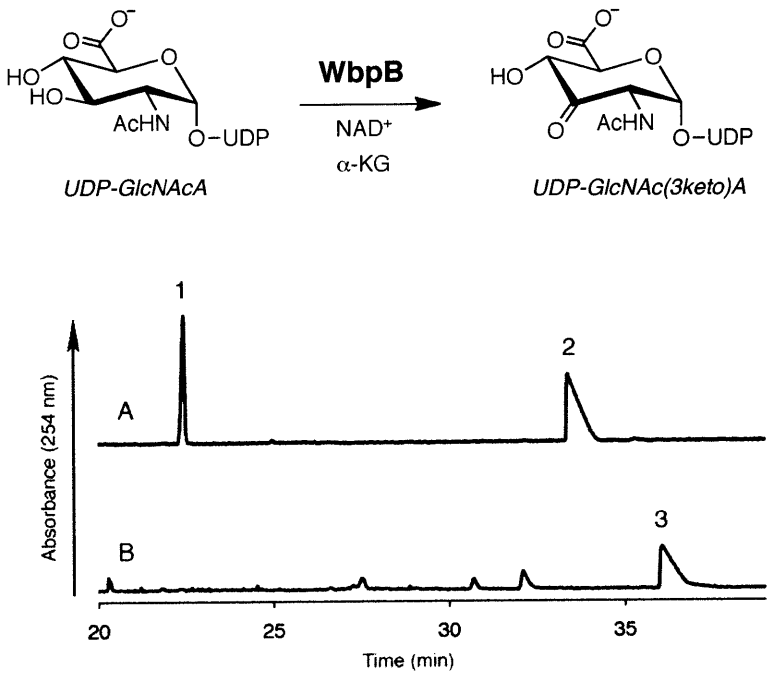


Figure 3-10: CE chromatogram representing (A) WbpB reaction in the absence of α-KG and WbpE, depicting no substrate conversion; (B) WbpB reaction containing 10 mM α-KG, depicting consumption of UDP-GlcNAcA. (1) NAD^+; (2) UDP-GlcNAcA; (3) UDP-GlcNAc(3keto)A.

To confirm that α-KG was in fact being consumed by WbpB, aliquots were removed from the reaction at specific time points and reacted with *o*-phenylenediamine (OPD), which has been previously shown to selectively label α-keto acids.[49,50] Analysis of the labeled aliquots by UV absorbance at 340 nm demonstrated a decrease in α-KG concentration over time, indicating that it was indeed being consumed (Figure 3-11). However, the use of a lower concentration of α-KG (2 mM) to maximize the observed signal decrease in the labeling reaction prevented complete turnover of the UDP-GlcNAcA starting material (70%). The product of α-KG reduction was identified as 2-hydroxyglutarate (2-HG) by thin-layer chromatography (TLC); development of TLC plates spotted with the crude reaction mixture followed by staining with

either dinitrophenylhydrazine, a marker of ketones and aldehydes, or bromocresol green, specific for carboxylates, revealed the presence of a newly formed product that did not contain a ketone and ran at the same R_f value as an authentic 2-HG standard.[51]

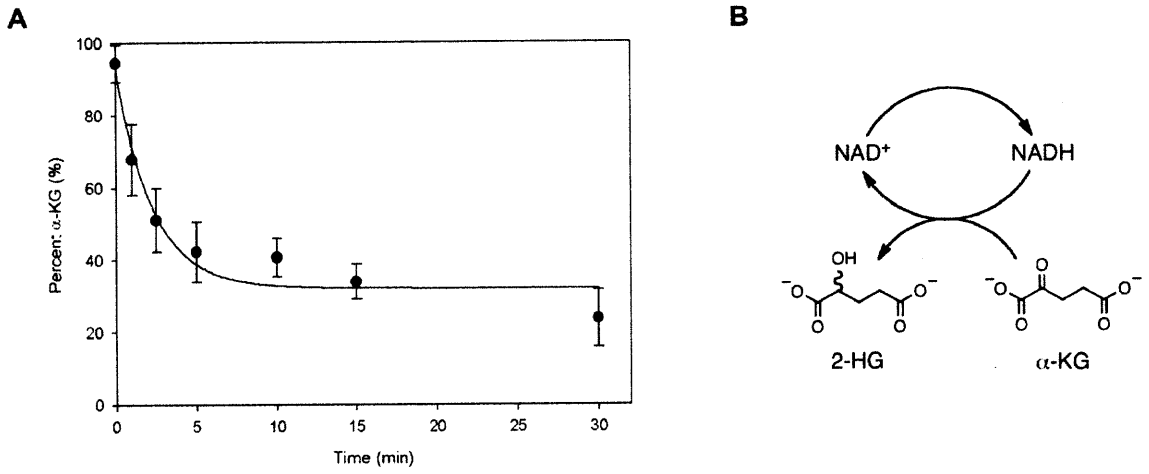


Figure 3-11: A) Verification of α-KG consumption over the course of the WbpB reaction, indicating it is the required oxidant for NAD^+ recycling. Aliquots were removed from the reaction at selected time points, quenched, labeled with OPD, and analyzed by absorbance at 340 nm. Data represent the average of three experiments; B) the reduction of α-KG to 2-HG by WbpB allows for efficient regeneration of NAD^+.

The necessary coupling of dehydrogenase/aminotransferase pairs in vitro has previously been reported; for example, the GnnA and GnnB enzymes from the Gram-negative acidophile *Acidithiobacillus ferrooxidans* must both be present to observe the conversion of UDP-GlcNAc to the corresponding C3"-modified UDP-GlcNAc($3NH_2$).[52] However, unlike GnnA and GnnB, the coupled WbpB/WbpE reaction does not require exogenous NAD^+ for product formation, supporting the claim that the cofactor is efficiently regenerated in situ over the course of the reaction. While the concomitant reduction of α-KG and oxidation of NADH has been previously observed in nature, such as the case of SerA 3-phosphoglycerate dehydrogenase,[53] to

our knowledge this is the first example of α-KG shunting in a nucleotide sugar biosynthesis pathway, thus representing a novel method of NAD^+ recycling.

A close comparison of the UDP-GlcNAc(3keto)A and α-KG structures offers insight into a possible pathway for recycling of NAD^+. As depicted in Figure 3-12, the terminal carboxylate, carbon backbone, and ketone of both molecules spatially overlap, suggesting similar binding interactions within the active site. One may envision a catalytic cycle in which initially, UDP-GlcNAcA is converted to UDP-GlcNAc(3keto)A, resulting in reduction of NAD^+ to NADH (Figure 3-12, 1). After dissociation of UDP-GlcNAc(3keto)A, a molecule of α-KG enters the active site (2) and is positioned favorably to permit formation of 2-HG and the oxidation of NADH back to NAD^+ (3), both completing the NAD^+ regeneration cycle and preparing WbpB for catalysis of another molecule of UDP-GlcNAcA.

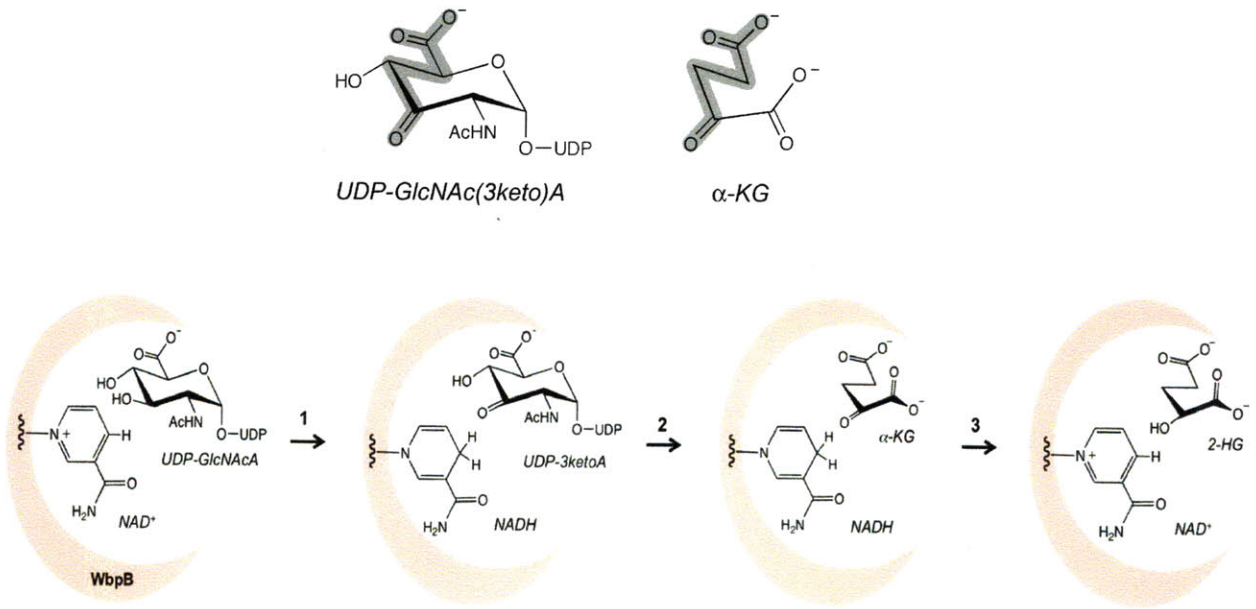


Figure 3-12: Comparison of UDP-GlcNAc(3keto)A and α-KG structures, suggesting similar binding orientations of the carboxylate and ketone moieties in the WbpB active site (*pink*).

The underlying necessity for the coupled activities of the dehydrogenase WbpB and aminotransferase WbpE remains unclear. One hypothesis is that UDP-GlcNAc(3keto)A is labile at physiological conditions, and thus the pairing of WbpB and WbpE activities allows for the direct transfer of the ketone intermediate from WbpB to WbpE. In support of this theory, the ketone has been shown to be unstable in water at room temperature, resulting in hydrolysis of the UDP moiety within several hours at pH 7. However, though the two enzymes may be in close proximity to deliver the ketone intermediate from WbpB to WbpE, the observation of WbpB function in the presence of α-KG implies that there is no required physical interaction between WbpB and WbpE. Future work in this area will entail evaluation of a putative WbpB/WbpE complex.

In August 2010, Thoden et al reported the crystal structure of WlbA, a close homolog of WbpB from *Thermus thermophilus* that also carries out the oxidation of UDP-GlcNAcA.[54] The authors present WlbA in complex with NAD^+ alone, as well as with either UDP-GlcNAcA or α-KG; interestingly, their structural analysis supports our claim that α-KG is the required oxidant for NAD^+ regeneration. As shown in Figure 3-13, both UDP-GlcNAcA and α-KG bind in the same place within the enzyme active site, with the C3" (UDP-GlcNAcA) and C2 (α-KG) carbons located within 3 Å of the *re* face of the nicotinamide cofactor. Analysis of the active site residues implicates Lys101 and His185 as playing important roles in catalysis, though more work is required to confirm this hypothesis. In addition, the authors compare the overall architecture of WlbA with that of the well-characterized sugar dehydrogenase, glucose-fructose oxidoreductase (GFOR, *Zymomonas mobilis*). Like WlbA, GFOR crystallized with a tetrameric quaternary structure; however, each protomer adopts a drastically different conformation compared with the other subunits in the complex (Figure 3-13C).[55] Also, the WlbA protein

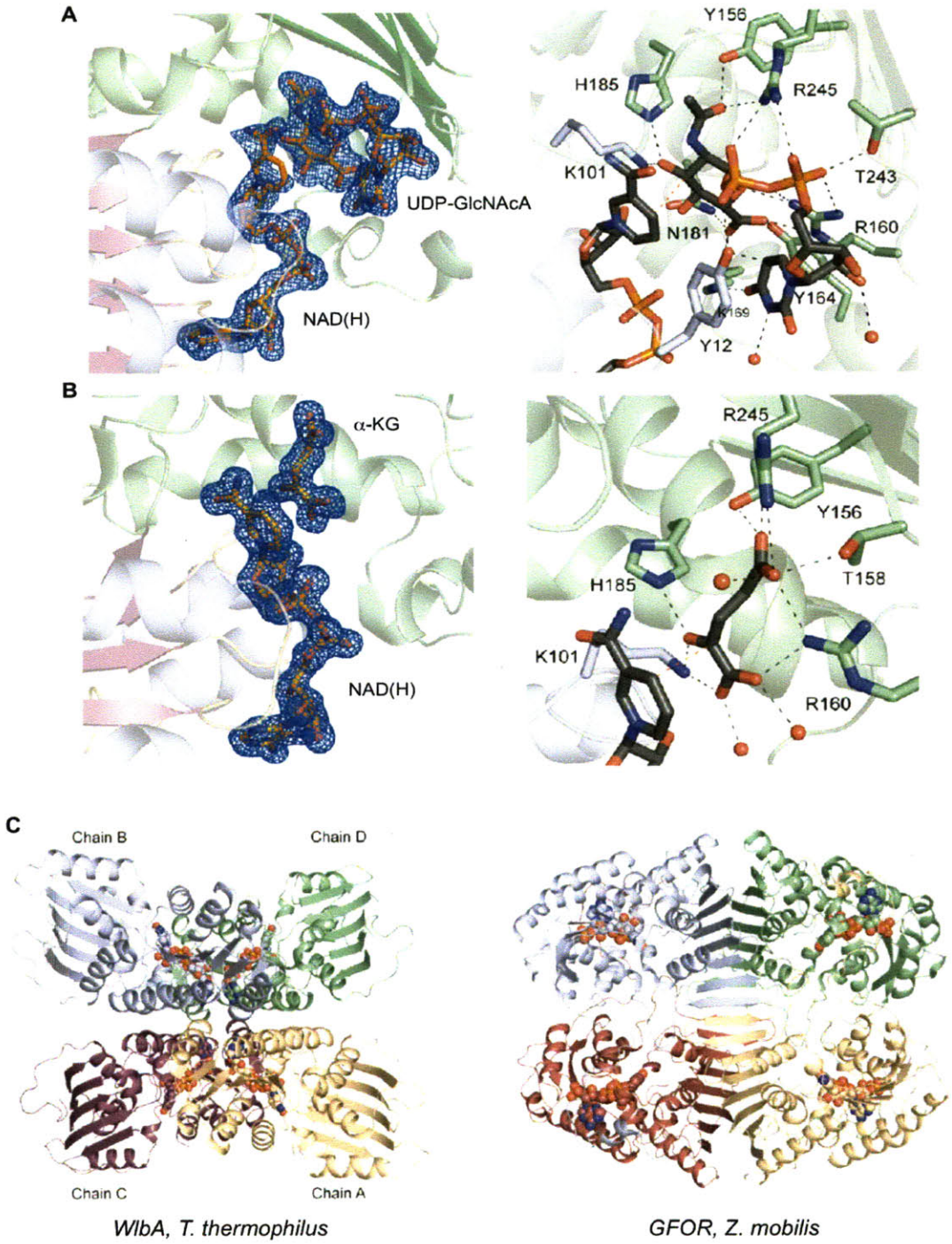


Figure 3-13: X-ray crystal structure of WlbA, a close WbpB homolog. WlbA bound to NAD(H) and UDP-GlcNAcA (A) or α-KG (B). A $2(F_o\text{-}F_c)$ electron density omit map contoured at 2.5 σ (*left*), and the active site residues located within 3.2 Å of either UDP-GlcNAcA or α-KG (*right*). Water molecules are depicted as red spheres and possible hydrogen bonds are shown as black lines. (C) Overall architecture of WlbA (*left*) and GFOR (*right*), highlighting the remarkable difference in protein structure. Figures are taken from Thoden et al.[54]

appears to bind its nicotinamide cofactor in an unprecedented manner, sharing it with two neighboring subunits. This unique structure helps to establish WbpB as one of the first examples of a new class of NAD^+ recycling nucleotide sugar dehydrogenases.

Functional Characterization of WbpD

The next enzyme in the UDP-ManNAc(3NAc)A biosynthetic pathway, WbpD, catalyzes the AcCoA-dependent acetylation of UDP-GlcNAc($3NH_2$)A to give UDP-GlcNAc(3NAc)A. Previous sequence analysis of WbpD indicated that it was predicted to have a left-handed β-helical (LβH) structure and thus is a member of the hexapeptide acyltransferase superfamily.[56,57] Previously characterized enzymes in this class of acyl- and acetyltransferases include the lipid A acyltransferase LpxA (*E. coli*), the glucosamine 1-phosphate *N*-acetyltransferase GlmU (*E. coli*), and the UDP-4-amino acetyltransferase PglD (*Campylobacter jejuni*), all of which contain numerous repeats of the hexapeptide motif L/I/V-[G/A/E/D]-X_4-[L/I/V] similar to WbpD.[58-60]

Incubation of WbpD with UDP-GlcNAc($3NH_2$)A and AcCoA resulted in the complete consumption of both UDP-GlcNAc($3NH_2$)A and AcCoA and the formation of two new peaks by CE (Figure 3-14). Using mass spectrometry and NMR analysis, these two peaks were identified to be CoA and UDP-GlcNAc(3NAc)A. The utilization of [^{3}H]-AcCoA in the place of AcCoA resulted in [^{3}H]-labeled UDP-GlcNAc(3NAc)A, which was confirmed after purification and analysis by scintillation counting. As in the case with the WbpB/WbpE coupled reaction, WbpD was able to catalyze the acetylation of UDP-GlcNAc($3NH_2$)A at a broad range of pH (5.5-10.0) and temperature (4-65 °C), although the optimal reaction conditions for catalysis were found to be 30 °C and pH 7.0 to limit base-promoted hydrolysis of AcCoA.

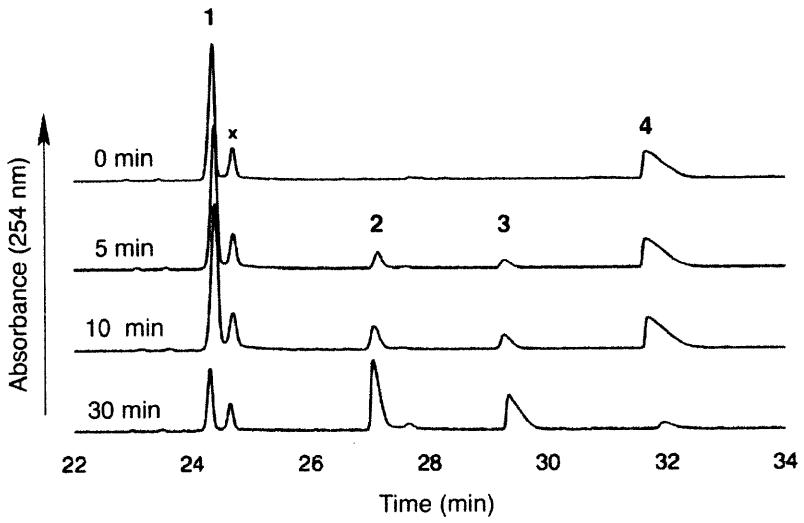


Figure 3-14: CE time course analysis of the WbpD reaction. (1) AcCoA; (2) CoA; (3) UDP-GlcNAc(3NAc)A; (4) UDP-GlcNAc(3NH_2)A. The peak labeled **x** represents an impurity present in the AcCoA starting material.

The specificity of WbpD for its nucleotide sugar substrate was investigated by incubation with UDP-GlcNAcA, UDP-GlcNAc, and UDP-2-acetamido-4-amino-2,4,6-trideoxy-D-glucosamine, the UDP-4-amino sugar product of the C4"-aminotransferase from *C. jejuni*, PglE.[61] No appreciable turnover of starting material was observed in any case, signifying the importance of an amine at the C3" position for acetylation. In addition to exploring substrate specificity, the kinetic parameters of the WbpD reaction were determined by varying the concentration of UDP-GlcNAc(3NH_2)A from 0.015-1.5 mM in the presence of fixed concentrations of AcCoA (1.5 mM) and WbpD (Figure 3-15). At the highest concentration of substrate, no more than 10% turnover was observed. Data were collected by manual integration of CE traces and then directly fit to the Michaelis-Menten equation using SigmaPlot (Systat Software). The conclusion from this analysis is that WbpD exhibits an apparent K_m = 107 ± 15

μM and k_{cat} = (2.9 ± 0.12) x 10^3 min^{-1} for its nucleotide sugar substrate, UDP-GlcNAc(3NH_2)A (Figure 3-15).

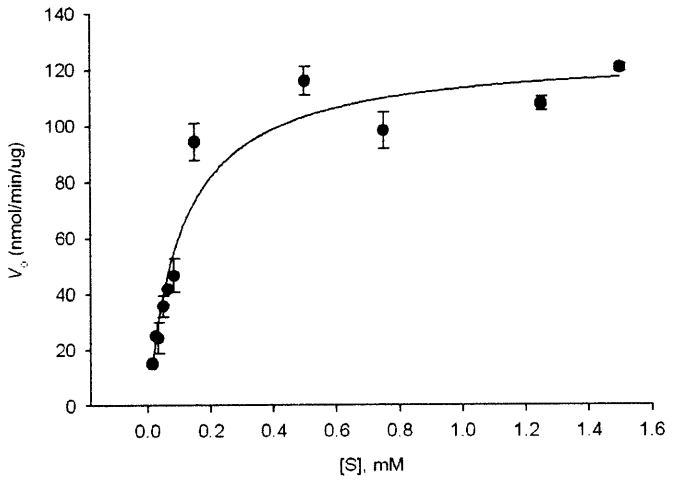


Figure 3-15: Michaelis-Menten diagram depicting WbpD kinetic parameters. Reactions were carried out with AcCoA as the fixed substrate (1.5 mM) and UDP-GlcNAc(3NH_2)A as the variable substrate (0.015-1.5 mM). Data are the average of two experiments.

Based on these kinetic parameters, it appears that WbpD has a greater affinity for its nucleotide sugar substrate UDP-GlcNAc(3NH_2)A (apparent K_m = 107 ± 16 μM) and catalyzes acetylation in a moderately rapid fashion (k_{cat} = 2.9 ± 0.12 x 10^3 min^{-1}) when compared with PglD, the C4" acetyltransferase involved in the biosynthesis of UDP-*N*,*N'*-diacetylbacillosamine (UDP-Bac) from *C. jejuni* characterized previously (K_m = 410 ± 78 μM and k_{cat} = 4.83 ± 0.30 x 10^5 min^{-1}).[62] PglD is a highly efficient enzyme (k_{cat}/K_m = 1.18 x 10^6 min^{-1} mM^{-1}), and it is hypothesized that its ability to rapidly carry out acetylation serves to prevent the buildup of early intermediates produced by the slower preceding enzymes (PglE and PglF) in the biosynthetic sequence. While the kinetic parameters of WbpB and WbpE have yet to be

determined, it is clear that WbpD (k_{cat}/K_m = 2.69 x 10^4 min^{-1} mM^{-1}) is in fact more efficient than WbpA (k_{cat}/K_m = 913 min^{-1} mM^{-1}), the first enzyme in the pathway, in turning over its nucleotide sugar substrate;[30] perhaps the increased efficiency of WbpD provides the same function in this case as well.

In early 2010, Thoden et al published the X-ray crystal structure of WlbB from *Bordetella petrii*, a nearly identical homolog of WbpD that also catalyzes the acetylation of UDP-GlcNAc(3NH_2)A.[63] The structure of this protein confirms that WlbB adopts a trimeric structure, in which each protomer is predominantly composed of a long, left-handed β-helix that represents the hallmark of the LβH hexapeptide acyltransferase superfamily (Figure 3-16A). The authors obtained a crystal structure of the complex between WlbB, CoA, and the UDP-GlcNAc(3NH_2)A substrate; surprisingly, they note that there are only two hydrogen bonds that can reasonably occur between the protein backbone and the nucleotide sugar, namely from Asn84 and Arg94. From this structure, it is unclear which atom or residue plays the role of the catalytic base in the reaction that would serve to deprotonate the C3" amine. Earlier work by Olivier et al finds that in the case of PglD, an active site histidine residue was found to behave as the catalytic base;[60] however, the structure of WlbA shows no histidine within 4 Å of the C3" amine capable of performing the deprotonation. In addition, the sulfur atom of CoA is ideally positioned in the active site to interact with the C3" amine of the nucleotide sugar (Figure 3-16C). For this reason, the authors propose a unique reaction mechanism, whereby the sulfur atom of AcCoA ultimately behaves as the catalytic base required to deprotonate the C3" amine in a concerted fashion; further evidence to support this hypothesis is still required.

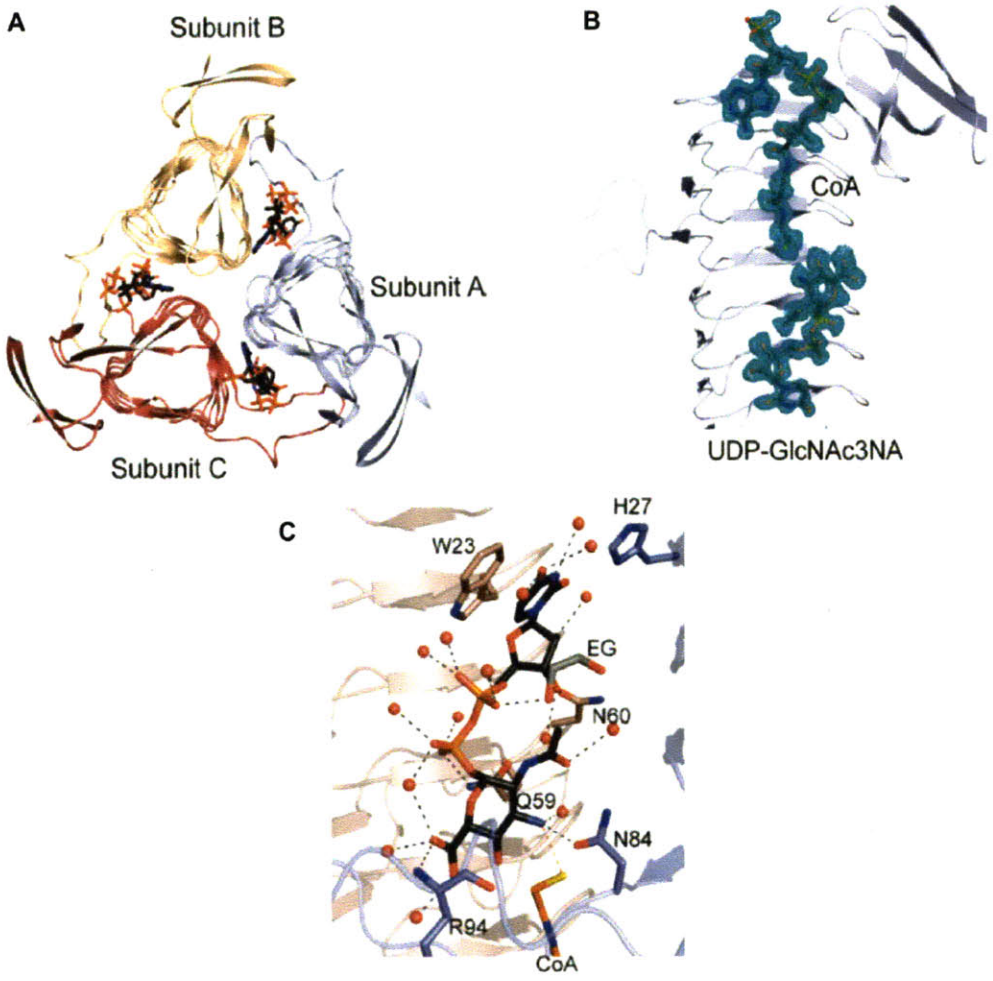


Figure 3-16: Crystal structure of WbpD homolog, WlbB, in complex with CoA and UDP-GlcNAc(3NH_2)A. (A) Ribbon representation of the WlbB trimer with the three subunits shown in red, beige and light blue. (B) A *2(F_o-F_c)* omit map contoured at 1.5σ, depicting the active site of a single protomer of WlbB. (C) Amino residues lying within 3.2 of UDP-GlcNAc(3NH_2)A. Water molecules are depicted as red spheres and possible hydrogen bonds are represented by dashed lines. Figures are taken from Thoden et al.[63]

One-Pot Synthesis of UDP-ManNAc(3NAc)A

As a final means to characterize the enzymes in the Wbp pathway, the synthesis of UDP-ManNAc(3NAc)A was attempted in a single reaction vessel using WbpA, WbpB, WbpE, WbpD and WbpI. WbpI, the C2" epimerase, was overexpressed and purified in the same manner as the

other four enzymes in the pathway. All five enzymes were combined with UDP-GlcNAc and the requisite cofactors; after incubation for 8 hours, CE analysis indicated partial consumption of the UDP-GlcNAc starting material (15%) and formation of a new peak (Figure 3-17). The crude reaction mixture was purified by RP-HPLC and the identity of UDP-ManNAc(3NAc)A was confirmed by high resolution ESI-MS, as well as comparison with a pure sample of UDP-GlcNAc(3NAc)A by CE. This result demonstrates that by combining WbpA, WbpB, WbpE, WbpD and WbpI with UDP-GlcNAc and obligate cofactors and substrates in vitro, it is possible to generate UDP-ManNAc(3NAc)A in a one-pot reaction. This multienzyme biotransformation now provides a platform for screening the entire sequence of Wbp enzymes for possible inhibitors in a single reaction vessel.

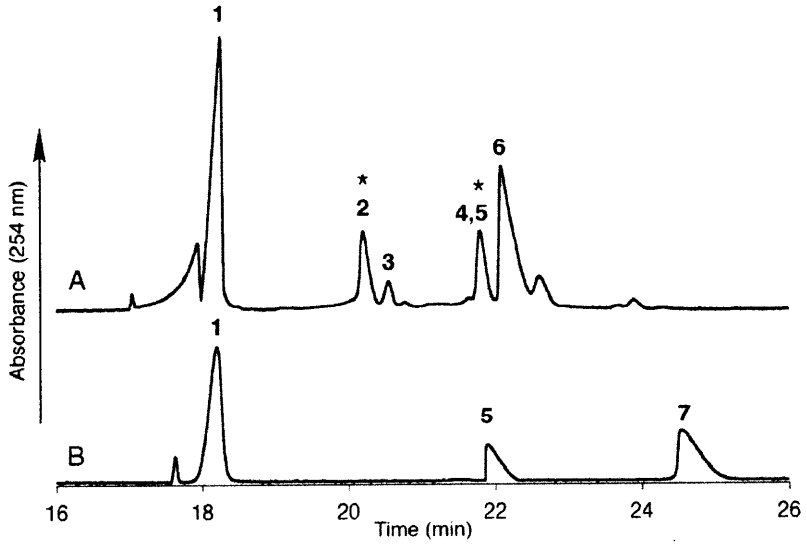


Figure 3-17: CE analysis of the (A) crude one-pot reaction of WbpA, WbpB, WbpE, WbpD and WbpI and (B) purified UDP-ManNAc(3NAc)A, with both NAD^+ and UDP-GlcNAc(3NAc)A included as references. UDP-ManNAc(3NAc)A has the same retention time as NADH. (1) NAD^+; (2) UDP-GlcNAc; (3) AcCoA; (4) NADH; (5) UDP-ManNAc(3NAc)A; (6) CoA; (7) UDP-GlcNAc(3NAc)A. The starred peaks refer to the starting material (2, UDP-GlcNAc) and the final product (5, UDP-ManNAc(3NAc)A).

Conclusions

In conclusion, this chapter describes the biochemical characterization of WbpB, WbpE, and WbpD, three key enzymes responsible for the formation of UDP-GlcNAc(3NAc)A. It provides a facile route to the chemoenzymatic synthesis of milligram quantities of this rare nucleotide sugar that can be used for studying *N*-linked glycosylation in archaea, and completes the annotation of the UDP-ManNAc(3NAc)A biosynthetic pathway in *P. aeruginosa* (Figure 3-18). Due to the presence of UDP-GlcNAc(3NAc)A and UDP-ManNAc(3NAc)A in related pathogens, such as *Bordetella pertussis*, *Bordetella bronchiseptica*, and other *Pseudomonas* strains, we envision that the methods outlined herein may provide useful tools to probe similar biosynthetic pathways in these organisms. Lastly, as these enzymes play a critical role in the formation of lipopolysaccharide, they may present new targets for the development of potential therapeutics for treatment of *P. aeruginosa* infection.

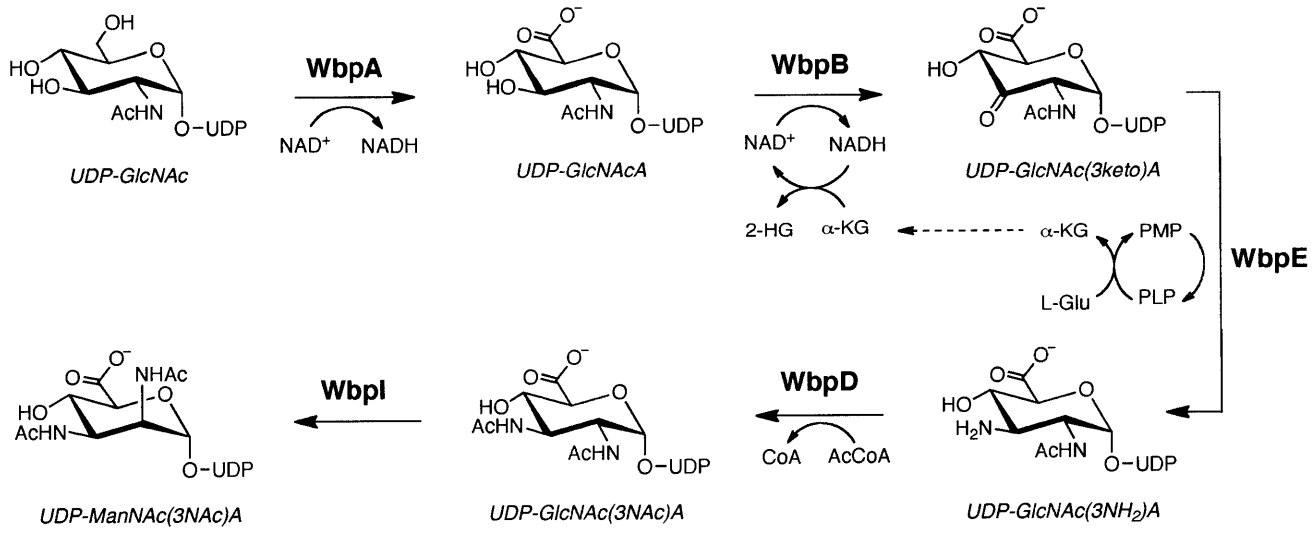


Figure 3-18: Biosynthetic pathway of UDP-ManNAc(3NAc)A in *P. aeruginosa* PAO1.

Acknowledgements

I am grateful to Dr. Jeff Simpson of the MIT Department of Chemistry Instrumentation Facility for assistance with NMR acquisition and Dr. Nelson Olivier for technical advice.

Experimental Methods

General Information

Unless otherwise noted, all reagents were obtained commercially and used without further purification. Oligonucleotides were purchased from Eurofins MWG Operon (Huntsville, AL) and Sigma Life Sciences (St. Louis, MO). Restriction endonucleases were obtained from New England Biolabs (Ipswich, MA). UDP-GlcNAc, UDP-GlcA, UDP-GalNAc, AcCoA, PLP, NAD^+ and $NADP^+$ and amino acids were purchased from Sigma Aldrich (St. Louis, MO). [^{3}H]AcCoA was acquired from American Radiolabeled Chemicals (St. Louis, MO), and ^{15}N-L-glutamate as well as deuterated solvents for NMR spectroscopy were obtained from Cambridge Isoptope Laboratories (Andover, MA). Shigemi tubes for NMR spectroscopy were purchased from Shigemi, Inc (Allison Park, PA). Analytical thin-layer chromatography (TLC) was carried out on silica gel 60 F254 plates (EMD Chemicals). Sequencing of all bacterial plasmids was performed by the MIT CCR Biopolymers Laboratory (Cambridge, MA). NMR spectra were acquired using either a Bruker 600 MHz spectrometer equipped with a 5 mm inverse cryoprobe, a Varian Inova 500 MHz spectrometer with a 5 mm inverse broadband gradient probe, or a Bruker 400 MHz spectrometer furnished with a 5 mm variable temperature gXH probe. Chemical shifts (δ) are reported in parts per million (ppm) and references to internal standards; the HOD signal at 4.80 ppm was used as for ^{1}H analysis and the $(CH_3\underline{C}H_2)_3N$ signal of

triethylamine at 47.19 ppm was utilized for ^{13}C observation. An external reference of 85% H_3PO_4 was employed for ^{31}P NMR. Coupling constants (J) are reported in Hertz (Hz) and multiplicites are abbreviated as singlet (s), doublet (d), triplet (t), multiplet (m), doublet of doublets (dd), and doublet of triplets (dt). The terms apparent triplet (t_{app}) and apparent quartet (q_{app}) are used to describe a doublet of doublets (dd) with two similar coupling constants or a doublet of doublet of doublets (ddd) with three similar coupling constants. High-resolution ESI-MS was performed by Department of Chemistry Instrumentation Facility at MIT (Cambridge, MA).

Cloning of wbpA, wbpB, wbpE, and wbpD

The genes were amplified by the polymerase chain reaction from *P. aeruginosa* PAO1-LAC genomic DNA (ATCC) using Pfu Turbo polymerase (Stratagene) and the oligonucleotides described in the Appendix (Table 1). The resulting PCR products containing both *Bam*HI and *Xho*I restriction sites were digested and cloned into the same sites of the pET24a(+) vector (Novagen) via standard molecular biology techniques. The final gene products encoded proteins with an N-terminal T7-tag and a C-terminal His_6-tag.

Overexpression of WbpA, WbpB, WbpE and WbpD

The pET24a(+) plasmids containing either *wbpA*, *wbpB*, *wbpE*, or *wbpD* were transformed into *E. coli* BL21-CodonPlus(DE3) RIL competent cells (Stratagene) using both kanamycin (50 μg/mL) and chloramphenicol (30 μg/mL) for selection. For overexpression, 1 L of Luria-Bertani media supplemented with kanamycin and chloramphenicol was inoculated at 37 °C while

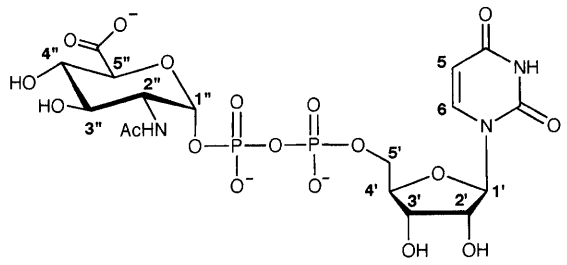
shaking until an optical density (600 nm) of 0.6-0.8 was achieved. The cultures were then cooled to 16 °C and protein expression was induced by the addition of IPTG (1 mM). After 16 hrs, the cells were harvested by centrifugation (5000 x *g*) and the resultant cell pellets were stored at -80 °C until needed.

Purification of WbpA, WbpB, WbpE, and WbpD

All steps were performed at 4 °C. The cell pellets from a 1 L culture were thawed and resuspended in 50 mL cold lysis buffer (50 mM HEPES, pH 8.0/300 mM NaCl/10 mM imidazole) with Protease Inhibitor Cocktail Set III (EDTA-free, Calbiochem) added at a 1000-fold dilution. The cells were disrupted by ultrasonication on ice, and the resulting cell lysate was then cleared of unbroken cells, cellular debris, and membranes by centrifugation (145,000 x *g*) for 65 minutes. The resulting supernatant was incubated with Ni-NTA agarose resin (Qiagen) for 2 hrs with gentle rocking and subsequently poured into a fritted PolyPrep column (BioRad) to collect the resin. The resin was washed (50 mM HEPES, pH 8.0/300 mM NaCl/25 mM imidazole) and the protein eluted (50 mM HEPES, pH 8.0/300 mM NaCl/250 mM imidazole). Fractions containing the desired product were combined and dialyzed (50 mM HEPES, pH 8.0/100 mM NaCl) to remove the imidazole and lower the salt concentration. Proteins were stored at either -20 °C (WbpA, WbpB, WbpE) or 4 °C (WbpD) after the removal of precipitate by filtration and addition of glycerol to a final concentration of 25%. Protein purity was measured by SDS-PAGE, and protein identity was confirmed by Western blot analysis using both Anti-T7 (Novagen) and Anti-His_4 (Qiagen) antibodies. Protein concentration was determined by either the Micro BCA kit (Pierce) or UV absorbance using the following

extinction coefficients at 280 nm: WbpA (32,320 $M^{-1}cm^{-1}$), WbpB (38,150 $M^{-1}cm^{-1}$), WbpE (20,040 $M^{-1}cm^{-1}$), WbpD (22,340 $M^{-1}cm^{-1}$).

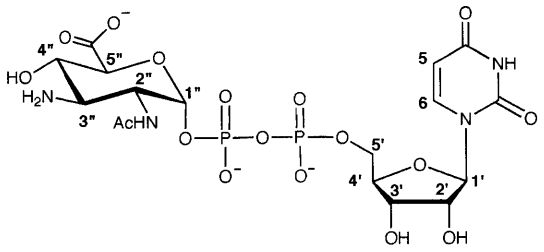
Chemical Synthesis of UDP-GlcNAcA



UDP-GlcNAcA was prepared via oxidation of UDP-GlcNAc as previously described.[37] Briefly, Adams' catalyst (10 mg) was activated in H_2O (1 mL) under H_2 for 2 hrs, then flushed with N_2. The catalyst was then transferred to a stirring mixture UDP-GlcNAc (100 mg, 0.153 mmol) and $NaHCO_3$ (25.8 mg, 0.307 mmol) in H_2O (5 mL) in a three-necked round bottom flask. A gas bubbler (medium grade) was inserted into the side neck, and the reaction was refluxed for 72 hrs in the presence of streaming O_2. Three additional aliquots of activated catalyst (10 mg) were added over the course of the reaction. After filtration to remove the Pt catalyst, the desired product was purified from the remaining UDP-GlcNAc starting material using a 5 mL HiTrap Q FF anion exchange column (GE Healthcare), eluting with a linear gradient of 0-0.5 M NH_4HCO_3 over 250 mL. Fractions containing UDP-GlcNAcA were combined, freeze-dried, and resuspended in H_2O for further purification using a Synergi C_{18} Hydro preparatory RP-HPLC column (4 μm, 80 Å, 250 x 21.2 mm, Phenomenex) for removal of impurities and salt. The HPLC column was equilibrated with 50 mM triethylammonium bicarbonate (TEAB, pH 7.1) and, after loading, the product was eluted using a gradient of 0-50% CH_3CN over 30 minutes.

Quantification of the product was carried out using the molar extinction coefficient of uridine (ε_{262nm} = 10,000 $M^{-1}cm^{-1}$). The UDP-sugar to be analyzed was dissolved in 150 μL D_2O and freeze-dried several times prior to NMR analysis using a Shigemi tube. The desired UDP-GlcNAcA was characterized by ^1H and ^{13}C NMR spectroscopy and ESI-MS (Appendix). ^1H NMR (600 MHz, D_2O): δ 2.07 (s, $C(O)CH_3$), 3.57 (t_{app} $J_{4",5"}$ = 10.1, H-4"), 3.78 (t_{app}, $J_{3",4"}$ = 10.0, H-3"), 3.97 (dt, $J_{2",3"}$ = 9.9, $J_{2",P}$ = 2.8, H-2"), 4.13 (d, H-5"), 4.16 (m, H-5'b), 4.19 (m, H-5'a), 4.23 (m, H-4'), 4.32 (m, H-2' and H-3'), 5.50 (dd, $J_{1",P}$ = 7.3, $J_{1",2"}$ = 2.8, H-1"), 5.91 (d, $J_{5,6}$ = 8.0, H-5), 5.95 (d, $J_{1',2'}$ = 4.4, H-1'), 7.95 (d, H-6); ^{13}C NMR (400 MHz, D_2O): δ 174 (C-6"), 172.5 ($C(O)CH_3$), 163 (C-4), 153 (C-2), 143 (C-6), 104 (C-5), 94 (C-1"), 86 (C-1'), 83 (C-4'), 75 (C-2'), 74 (C-5"), 72 (C-4"), 71 (C-3"), 70 (C-3'), 65 (C-5'), 53 (C-2"), 22 ($C(O)CH_3$). HRMS: calcd for protonated acid $C_{17}H_{24}N_3O_{18}P_2$ ($[M]^{3-}$): requires *m/z* = 206.0130, found *m/z* = 206.0128 (ESI). NMR data were found to be in good agreement with previously reported values.[37]

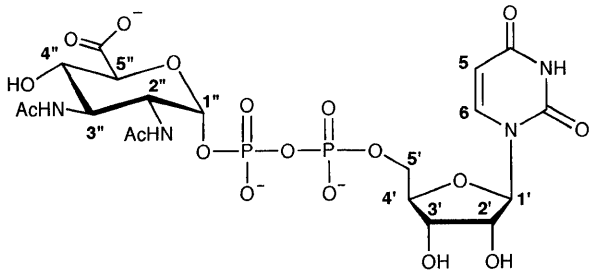
Synthesis of UDP-GlcNAc(3NH$_2$)A by the Coupled Reaction of WbpB and WbpE



WbpB/WbpE coupled reactions to produce UDP-GlcNAc(3NH_2)A contained 4.5 mg of each enzyme, 0.75 mM UDP-GlcNAcA, 0.2 mM NAD^+, 25 mM L-glutamate, 0.1 mM PLP, 50 mM HEPES (pH 8.0), 2.5 mM DTT, and 2 mM $MgCl_2$ in a total volume of 50 mL. Reactions were

incubated at 30 °C for 24 hours, during which time the progress of the reaction was monitored by CE. Protein was removed from the mixture by filtration and the resulting filtrate was frozen and lyophilized. The crude mixture was resuspended in H_2O and loaded onto a Synergi C_{18} Hydro preparatory RP-HPLC column equilibrated with 50 mM TEAB (pH 7.1). The UDP-GlcNAc($3NH_2$)A product was eluted with a linear gradient of 0-50% CH_3CN over 65 minutes. For ^{15}N-labeling of UDP-GlcNAc($3NH_2$)A, ^{15}N-L-glutamate was utilized in the place of L-glutamate. For determination of optimal pH, reactions included 50 mM MES (pH 5.5, 6.0, 6.5), 50 mM HEPES (pH 7.0, 7.5 8.0), 50 mM Bicine (pH 8.5), or 50 mM CHES (pH 9.0, 9.5, 10.0). To study temperature requirements, reactions were incubated at 4, 16, 25, 30, 37, 42, and 65 °C for the appropriate time. In all cases, the presence of product was confirmed by the addition of starting material into reaction mixtures and the observation of a new peak by CE. The desired UDP-GlcNAcA was characterized by ^1H and ^{13}C NMR spectroscopy and ESI-MS (Appendix). ^1H NMR (600 MHz, D_2O): δ 2.09 (s, C(O)CH_3), 3.57 (t_{app}, $J_{3",4"}$ = 10.0, H-3"), 3.80 (t_{app}, $J_{4",5"}$ = 10.1, H-4"), 4.17 (q_{app}, $J_{5'a,5'b}$ = 11.7, H-5'a), 4.19 (m, H-5"), 4.23 (m, H-5'b), 4.29 (m, H-4'), 4.32 (t_{app}, $J_{2",3"}$ = 11.0, H-2"), 4.36 (m, H-2' and H-3'), 5.57 (dd, $J_{1",P}$ = 7.4, $J_{1",2"}$ = 3.2, H-1"), 5.95 (d, $J_{5,6}$ = 8.1, H-5), 5.97 (d, $J_{1',2'}$ = 4.4, H-1'), 7.94 (d, H-6). ^{13}C NMR (400 MHz, D_2O): δ 174 (C-6"), 172.5 (C(O)CH_3), 163 (C-4), 153 (C-2), 143 (C-6), 104 (C-5), 95 (C-1"), 86 (C-1'), 83 (C-4'), 75 (C-2'), 74 (C-5"), 73 (C-4"), 70 (C-3'), 65 (C-5'), 54 (C-2"), 53 (C-3"), 23 (C(O)CH_3). HRMS: calcd for protonated acid $C_{17}H_{24}N_4O_{17}P_2$ ($[M^{2-}+H]^-$): requires *m/z* = 619.0695, found *m/z* = 619.0686 (ESI); for ^{15}N-incorporation, calcd for protonated acid $C_{17}H_{24}N_3{}^{15}NO_{17}P_2$ [M^-]: requires *m/z* = 620.0666, found *m/z* = 620.0687 (ESI).

Acetylation of UDP-GlcNAc(3H$_2$)A by WbpD



For synthesis of UDP-GlcNAc(3NAc)A by WbpD, reactions contained 0.75 mM UDP-GlcNAc(3NH$_2$)A, 0.75 mM AcCoA and 50 mM HEPES (pH 7.0) in a final volume of 7 mL. The reactions were incubated at 30 °C for 2 hours prior to CE analysis. The crude reaction mixture was filtered, lyophilized, and purified as described for UDP-GlcNAc(3NH$_2$)A. To incorporate [^{3}H] into UDP-GlcNAc(3NAc)A, [^{3}H]-AcCoA was used in the place of AcCoA. Determination of optimal pH and temperature was carried out as described above for the coupled WbpB/WbpE reaction. The desired UDP-GlcNAcA was characterized by ^{1}H and ^{13}C NMR spectroscopy and ESI-MS (Appendix). ^{1}H NMR (600 MHz, D_2O): δ 1.97 (s, C(O)CH_3), 2.00 (s, C(O)CH_3), 3.64 (t_{app}, $J_{4",5"}$ = 9.8, H-4"), 4.13 (t_{app}, $J_{2",3"}$ = 10.1, H-2"), 4.16 (t_{app} $J_{3",4"}$ = 9.8, H-3"), 4.19 (m, H-5'a), 4.24 (m, H-5'b and H-5"), 4.29 (m, H-4'), 4.37 (m, H-2' and H-3'), 5.54 (dd, $J_{1",P}$ = 7.1, $J_{1",2"}$ = 2.7, H-1"), 5.96 (d, $J_{5,6}$ = 8.1, H-5), 5.99 (d, $J_{1',2'}$ = 4.4, H-1'), 7.95 (d, H-6). ^{13}C NMR (400 MHz, D_2O): δ 176.0 (C-6"), 175.6 (*C*(O)CH_3), 175.4 (*C*(O)CH_3), 167.1 (C-4), 152.7 (C-2), 142.5 (C-6), 103.5 (C-5), 94.6 (C-1"), 89.0 (C-1'), 84.2 (C-4'), 74.6 (C-2'), 74.2 (C-5"), 70.9 (C-4"), 70.6 (C-3'), 65.9 (C-5'), 59.5 (C-3"), 53.0 (C-2"), 22.9 (C(O)CH_3), 22.7 (C(O)CH_3). HRMS: calcd for protonated acid $C_{19}H_{25}N_4O_{18}P_2$ ([M^{3-}+2H]$^-$): requires *m/z* = 661.0801, found *m/z* = 661.0790 (ESI).

Analysis of Reaction Products by Capillary Electrophoresis

CE was performed using a P/ACE MDQ system (Beckman Coulter) equipped with a UV detector. Bare silica capillary (75 μm x 80 cm) was utilized with detection at 72 cm and 25 mM sodium tetraborate (pH 9.5) as the running buffer. Prior to each run, the capillary was conditioned sequentially with 0.1 M NaOH, H_2O, and running buffer for 2 minutes. Samples were introduced to the capillary by pressure injection for 15 seconds at 30 mbar and separation was performed at 22 kV and monitored by UV absorbance at 254 nm. In general, samples were prepared by filtration with a 5K MWCO membrane (Millipore) and diluted (2x) with H_2O. Manual peak integration was carried out using Beckman 32 Karat software suite.

Verification of α-Ketoglutarate as the Required Oxidant for WbpB Function

To identify if α-KG was the required oxidant for NAD^+ regeneration by WbpB, a reaction containing 0.75 mM UDP-GlcNAcA, 50 mM HEPES (pH 8.0), 10 mM α-KG, and 5 μg WbpB in a total volume of 60 uL was assembled. The reaction was incubated at 30 °C for two hours prior to analysis by CE. After removal of the enzyme by filtration, thin-layer chromatography was utilized to observe the formation of 2-hydroxyglutarate (2-HG) in the crude reaction mixture. Samples were run on silica gel 60 F_{254} plates in a butanol/formic acid/H_2O mixture (8:3:2), air-dried, and developed either directly with dinitrophenylhydrazine or heated at 120 °C and stained with bromocresol green.[51]

In order to confirm the consumption of α-KG by WbpB, a similar reaction containing 0.75 mM UDP-GlcNAcA, 50 mM HEPES (pH 8.0), 2 mM α-KG and 2.5 μg WbpB in a volume of 250 μL was prepared. Aliquots (25 μL) were removed from the reaction at several time points,

quenched in 225 μL 1 M H_3PO_4, and split into two 125 μL fractions. A solution of *o*-phenylenediamine (OPD, 30 μL), a specific labeling agent for α-keto acids, [49,50] was then added to one 125 μL fraction, while the other was kept as an unlabeled control. The freshly prepared OPD solution consisted of 1 mg OPD in 1 mL H_3PO_4 (adjusted to pH 2 with 1 M NaOH) and 2.5 μL β-mercaptoethanol. All OPD-labeled samples were boiled for 5 minutes and then cooled to room temperature prior to analysis by UV absorbance at 340 nm. The absorbance measurements of the unlabeled fractions were subtracted from readings of the corresponding OPD-labeled samples; the resulting data represent the average of three experiments.

Identification of the Cofactor Bound to WbpB

In order to identify whether WbpB utilizes NAD^+ or $NADP^+$ as the preferred tightly bound cofactor, procedures previously outlined were employed with slight modification.[45,46] Briefly, 500 μg of purified WbpB (in 2 mL 50 mM HEPES, pH 8.0/100 mM NaCl) was precipitated by incubation with 10 μL 60% aqueous $HClO_4$ on ice for 20 minutes followed by centrifugation (7,000 x *g*). The supernatant was removed and the precipitate washed twice with 250 mL cold HEPES/NaCl buffer and spun down. The combined supernatants were neutralized with saturated $NaHCO_3$ and freeze-dried. The dried extract was then resuspended in 300 μL alcohol dehydrogenase buffer (45 mM glycine/75 mM sodium pyrophosphate, pH 9.0/170 mM EtOH) and transferred to a cuvette. NAD^+-specific alcohol dehydrogenase (10 μg, 445 units/mg, Sigma-Aldrich) was added and NADH formation was monitored by UV absorbance at 340 nm. After 10 minutes, $NADP^+$ (20 μM) was added to assess cofactor specificity and 5 minutes later, NAD^+ (20 μM) was introduced as a control for enzyme activity. The same procedure was performed using the $NADP^+$-specific isocitric dehydrogenase (20 μg, 95 units/mg, Sigma-

Aldrich), using a solution containing 6 mM isocitric acid and 50 mM HEPES (pH 7.0) to dissolve the WbpB extract.[46]

Determination of Kinetic Parameters of WbpD

For determination of UDP-GlcNAc(3NH_2)A kinetics, all reactions contained 0.5 ng of freshly purified WbpD, 1.5 mM AcCoA, 50 mM HEPES (pH 7.0), 5 μg bovine serum albumin as a carrier protein, and varying concentrations of UDP-GlcNAc(3NH_2)A (0.015-1.5 mM) in a total reaction volume of 30 μL. The reactions were incubated at 30 °C for 45 minutes, boiled for 2 minutes to inactivate the enzyme, and analyzed by CE to determine the amount of UDP-GlcNAc(3NAc)A produced. The kinetic parameters were determined from linear regression analysis and are the average of two experiments.

One-Pot Synthesis of UDP-ManNAc(3NAc)A

In order to synthesize UDP-ManNAc(3NAc)A in a one-pot process, 50 μg each of WbpA, WbpB, WbpE, WbpD and WbpI was incubated with 0.75 mM UDP-GlcNAc, 2.0 mM NAD^+, 100 mM $(NH_4)_2SO_4$, 25 mM L-glutamate, 0.1 mM PLP, 0.75 mM AcCoA, 2.5 mM DTT, 2 mM $MgCl_2$, and 50 mM HEPES (pH 7.0) in a total reaction volume of 2 mL. The reaction was incubated at 30 °C for 8 hours, filtered, and purified by RP-HPLC. Formation of the desired product was confirmed by high resolution ESI-MS and the observation of a distinct peak by CE when compared to a sample of UDP-GlcNAc(3NAc)A. WbpI was previously characterized and cloned, overexpressed, and purified as described above for WbpA, WbpB, WbpE, and WbpD.[64]

References

1. Pier, G.B. *Pseudomonas aeruginosa* lipopolysaccharide: A major virulence factor, initiator of inflammation and target for effective immunity. *Int. J. Med. Microbiol.* **2007**, *297*, 277-295.
2. Neu, H.C. Infections due to gram-negative bacteria: An overview. *Rev. Infect. Dis.* **1985**, *7*, S778-S782.
3. Diaz, M.H.; Shaver, C.M.; King, J.D.; Musunuri, S.; Kazzaz, J.A.; Hauser, A.R. *Pseudomonas aeruginosa* induces localized immunosuppression during pneumonia. *Infect. Immun.* **2008**, *76*, 4414-4421.
4. Tang, H.B.; DiMango, E.; Bryan, R.; Gambello, M.; Iglewski, B.H.; Goldberg, J.B.; Prince, A. Contribution of specific *Pseudomonas aeruginosa* virulence factors to pathogenesis of pneumonia in a neonatal mouse model of infection. *Infect. Immun.* **1996**, *64*, 37-43.
5. Son, M.S.; Matthews, W.J., Jr.; Kang, Y.; Nguyen, D.T.; Hoang, T.T. In vivo evidence of *Pseudomonas aeruginosa* nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. *Infect. Immun.* **2007**, *75*, 5313-5324.
6. Hancock, R.E. Resistance mechanisms in *Pseudomonas aeruginosa* and other nonfermative gram-negative bacteria. *Clin. Infect. Dis.* **1998**, *27*, S93-S99.
7. Mesaros, N.; Nordmann, P.; Plesiat, P.; Roussel-Delvallez, M.; Van Eldere, J.; Y., G.; Van Laethem, Y.; Jacobs, F.; Lebecque, P.; Malfroot, A.; Tulkens, P.M.; Van Bambeke, F. *Pseudomonas aeruginosa*: resistance and therapeutic options at the turn of the new millennium. *Clin. Microbiol. Infect.* **2007**, *13*, 560-578.
8. Zaborina, O.; Holbrook, C.; Chen, Y.; Long, J.; Zaborin, A.; Morozova, I.; Fernandez, H.; Wang, Y.; Turner, J.R.; Alverdy, J.C. Structure-function aspects of PstS in multi-drug resistant *Pseudomonas aeruginosa*. *PLoS Pathog.* **2008**, *4*, e43.
9. Döring, G.; Pier, G.B. Vaccines and immunotherapy against *Pseudomonas aeruginosa*. *Vaccine* **2008**, *26*, 1011-1024.
10. Cryz, S.J., Jr.; Pitt, T.L.; Furer, E.; Germanier, R. Role of lipopolysaccharide in virulence of *Pseudomonas aeruginosa*. *Infect. Immun.* **1984**, *44*, 508-513.
11. Lam, M.Y.; McGroarty, E.J.; Kropinski, A.M.; MacDonald, L.A.; Pedersen, S.S.; Hoiby, N.; Lam, J.S. Occurrence of a common lipopolysaccharide antigen in standard and clinical strains of *Pseudomonas aeruginosa*. *J. Clin. Microbiol.* **1989**, *27*, 962-967.
12. Rocchetta, H.L.; Burrows, L.L.; Lam, J.S. Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 523-553.
13. Knirel, Y.A.; Bystrova, O.V.; Kocharova, N.A.; Zahringer, U.; Pier, G.B. Conserved and variable structural features in the lipopolysaccharide of *Pseudomonas aeruginosa*. *J. Endotoxin Res.* **2006**, *12*, 324-336.
14. Arsenault, T.L.; Hughes, D.W.; MacLean, D.B.; Szarek, W.; Kropinkski, A.M.B.; Lam, J.S. Structural studies on the polysaccharide portion of "A-band" lipopolysaccharide from a mutant (AK1401) of *Pseudomonas aeruginosa* strain PAO1. *Can. J. Chem.* **1991**, *69*, 1273-1280.
15. Knirel, Y.A.; Vinogradov, E.V.; Kocharova, N.A.; Paramonov, N.K.; Kochetkov, N.K.; Dmitriev, B.A.; Stanislavsky, E.S.; Lanyi, B. The structure of O-specific polysaccharides and serological classification of *Pseudomonas aeruginosa*. *Acta Microbiol. Hung.* **1988**, *35*, 3-24.

16. Bystrova, O.V.; Knirel, Y.A.; Lindner, B.; Kocharova, N.A.; Kondakova, A.N.; Zahringer, U.; Pier, G.B. Structures of the core oligosaccharide and O-units in the R- and SR-type lipopolysaccharides of reference strains of *Pseudomonas aeruginosa* O-serogroups. *FEMS Immunol. Med. Microbiol.* **2006**, *46*, 85-99.
17. Dasgupta, T.; de Kievit, T.R.; Masoud, H.; Altman, E.; Richards, J.C.; Sadovskaya, I.; Speert, D.P.; Lam, J.S. Characterization of lipopolysaccharide-deficient mutants of *Pseudomonas aeruginosa* derived from serotypes O3, O5, and O6. *Infect. Immun.* **1994**, *62*, 809-817.
18. Hancock, R.E.; Mutharia, L.M.; Chan, L.; Darveau, R.P.; Speert, D.P.; Pier, G.B. *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis: a class of serum-sensitive, nontypable strains deficient in lipopolysaccharide O side chains. *Infect. Immun.* **1983**, *42*, 170-177.
19. Kochetkov, N.K.; Knirel, Y.A. Structure of lipopolysaccharides from gram-negative bacteria. III. Structure of O-specific polysaccharides. *Biochemistry (Moscow)* **1994**, *59*, 1325-1383.
20. Burrows, L.L.; Charter, D.F.; Lam, J.S. Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PAO1) B-band lipopolysaccharide gene cluster. *Mol. Microbiol.* **1996**, *22*, 481-495.
21. Kintz, E.; Goldberg, J.B. Regulation of lipopolysaccharide O-antigen expression in *Pseudomonas aeruginosa*. *Future Microbiol* **2008**, *3*, 191-203.
22. Rocchetta, H.L.; Burrows, L.L.; Pacan, J.C.; Lam, J.S. Three rhamnosyltransferases responsible for assembly of the A-band D-rhamnan polysaccharide in *Psuedomonas aeruginosa*: a fourth transferase, WbpL, is required for the initiation of both A-band and B-band lipopolysaccharide synthesis. *Mol. Microbiol.* **1998**, *28*, 1103-1109.
23. Whitfield, C. Biosynthesis of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* **2006**, *75*, 39-68.
24. Raetz, C.R.H.; Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **2002**, *71*, 635-700.
25. Daniels, C.; Griffiths, C.; Cowles, B.; Lam, J.S. Pseudomonas aeruginosa O-antigen chain length is determined before ligation to lipid A core. *Environ Microbiol* **2002**, *4*, 883-897.
26. Franco, A.V.; Liu, D.; Reeves, P.R. The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. *J. Bacteriol.* **1998**, *180*, 2670-2675.
27. Larue, K.; Kimber, M.S.; Ford, R.; Whitfield, C. Biochemical and structural analysis of bacterial O-antigen chain length regulator proteins reveals a conserved quaternary structure. *J. Biol. Chem.* **2009**, *280*, 7395-7403.
28. Papadopoulos, M.; Morona, R. Mutagenesis and chemical cross-linking suggest that Wzz dimer stabilty and oligomerization affect lipopolysaccharide O-antigen modal chain length control. *J. Bacteriol.* **2010**, *192*, 3385-3393.
29. Chng, S.-S.; Ruiz, N.; Chimalakonda, G.; Silhavy, T.J.; Kahne, D. Characterization of the two-protein complex in *Escherichia coli* responsible for the lipopolysaccharide assembly at the outer membrane. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5363-5368.
30. Miller, W.L.; Wenzel, C.Q.; Daniels, C.; Larocque, S.; Brisson, J.-R.; Lam, J.S. Biochemical characterization of WbpA, a UDP-*N*-acetyl-D-glucosamine 6-

dehydrogenase involved in O-antigen biosynthesis in *Pseudomonas aeruginosa* PAO1. *J. Biol. Chem.* **2004**, *279*, 37551-37558.

31. Voisin, S.; Houliston, R.S.; Kelly, J.; Brisson, J.-R.; Watson, D.; Bardy, S.L.; Jarrell, K.F.; Logan, S.M. Identification and characterization of the unique *N*-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *J. Biol. Chem.* **2005**, *280*, 16586-16593.
32. Kelly, J.; Logan, S.M.; Jarrell, K.F.; VanDyke, D.J.; Vinogradov, E.V. A novel *N*-linked flagellar glycan from *Methanococcus maripaludis*. *Carb. Res.* **2009**, *344*, 648-653.
33. Rejzek, M.; Kannathasan, V.S.; Wing, C.; Preston, A.; Westman, E.L.; Lam, J.S.; Naismith, J.H.; Maskell, D.J.; Field, R.A. Chemical synthesis of UDP-Glc-2,3-diNAcA, a key intermediate in cell surface polysaccharide biosynthesis in the human respiratory pathogens *B. pertussis* and *P. aeruginosa*. *Org. Biomol. Chem.* **2009**, *7*, 1203-1210.
34. Fan, D.-F.; John, C.E.; Zalitis, J.; Feingold, D.S. UDP-acetylglucosamine dehydrogenase from *Achromobacter georgiopolitanum*. *Arch. Biochem. Biophys.* **1969**, *135*, 45-49.
35. Kawamura, T.; Ichihara, N.; Sugiyama, S.; Yokota, H.; Ishimotot, N.; Ito, E. Biosynthesis of UDP-*N*-acetyl-D-glucosaminuronic acid and UDP-*N*-acetyl-D-mannosaminuronic acid in *Micrococcus luteus*. *J. Biochem.* **1985**, *98*, 105-116.
36. Blom, N.S.; Tétreault, S.; Coulombe, R.; Sygusch, J. Novel active site in *Eschericha coli* fructose 1,6-bisphosphate aldolase. *Nat. Struct. Biol.* **1996**, *3*, 856-862.
37. Rejzek, M.; Mukhopadhyay, B.; Wenzel, C.Q.; Lam, J.S.; Field, R.A. Direct oxidation of sugar nucleotides to the corresponding uronic acids: TEMPO and platinum-based procedures. *Carb. Res.* **2007**, *342*, 460-466.
38. Marino, J.P.; Fernandez de la Pradilla, R.; Laborde, E. Regio- and stereoselectivity of the reaction between cyanocuprates and cyclopentene epoxides. Application to the total synthesis of prostaglandins. *J. Org. Chem.* **1987**, *52*, 4898-4913.
39. Kim, K.-H.; Kim, J.-Y.; Lee, K.-H.; Noh, M.-J.; Kim, Y.-C.; Park, H.-J. Synthesis and biological activity of the new 5-fluorocytosine derivatives, 5'-deoxy-*N*-alkyloxycarbonyl-5-fluorocytosine-5'-carboxylic acid. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 483-486.
40. Rao, S.T.; Rossmann, M.G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **1973**, *76*, 241-256.
41. Burgie, E.S.; Thoden, J.B.; Holden, H.M. Molecular architecture of DesV from *Streptomyces venezuelae*: A PLP-dependent transaminase involved in the biosynthesis of the unusual sugar desosamine. *Protein Sci* **2007**, *16*, 887-896.
42. Schoenhofen, I.C.; Lunin, V.V.; Julien, J.-P.; Li, Y.; Ajamian, E.; Matte, A.; Cygler, M.; Brisson, J.-R.; Aubry, A.; Logan, S.M.; Bhatia, S.; Wakarchuk, W.W.; Young, N.M. Structural and functional characterization of PseC, an aminotransferase involved in the biosynthesis of pseudaminic acid, an essential flagellar modification in *Helicobacter pylori*. *J. Biol. Chem.* **2006**, *281*, 8907-8916.
43. Obhi, R.K.; Creuzenet, C. Biochemical characterization of the *Campylobacter jejuni* Cj1294, a novel UDP-4-keto-6-deoxy-GlcNAc aminotransferase that generates UDP-4-amino-4,6-dideoxy-GalNAc. *J. Biol. Chem.* **2005**, *280*, 20902-20908.
44. Vijayakumar, S.; Merkx-Jacques, A.; Ratnayake, D.B.; Gryski, I.; Obhi, R.K.; Houle, S.; Dozois, C.M.; Creuzenet, C. Cj1121c, a novel UDP-4-keto-6-deoxy-GlcNAc C-4 aminotransferase essential for protein glycosylation and virulence in *Campylobacter jejuni*. *J. Biol. Chem.* **2006**, *281*, 27733-27743.

45. Klepp, J.; Retey, J. The stoichiometry of the tightly bound NAD^+ in urocanase: Separation and characterization of fully active and inhibited forms of the enzyme. *Eur. J. Biochem.* **1989**, *185*, 615-619.
46. Ni, Y.; McPhie, P.; Deacon, A.; Ealick, S.; Coleman, W.G., Jr. Evidence that $NADP^+$ is the physiological cofactor of ADP-L-glycero-D-mannoheptose 6-epimerase. *J. Biol. Chem.* **2001**, *276*, 27329-27334.
47. Hegeman, A.D.; Gross, J.W.; Frey, P.A. Concerted and stepwise dehydration mechanisms observed in the wild-type and mutated *Escherichia coli* dTDP-glucose 4,6-dehydratase. *Biochemistry* **2002**, *41*, 2797-2804.
48. Morrison, J.P.; Schoenhofen, I.C.; Tanner, M.E. Mechanistic studies on PseB of pseudaminic acid biosynthesis: A UDP-N-acetylglucosamine 5-inverting 4,6-dehydratase. *Bioorg Chem* **2008**, *36*, 312-320.
49. Lanning, M.C.; Cohen, S.S. The detection and estimation of 2-ketohexonic acids. *J. Biol. Chem.* **1951**, *189*, 109-114.
50. Kalliri, E.; Mulrooney, S.B.; Hausinger, R.P. Identification of *Escherichia coli* YgaF as an L-2-hydroxyglutarate oxidase. *J. Bacteriol.* **2008**, *190*, 3793-3798.
51. Stahl, E. *Thin-Layer Chromatography: A Laboratory Handbook*; 2nd Ed.; Springer-Verlag: New York, 1969.
52. Sweet, C.R.; Ribeiro, A.A.; Raetz, C.R.H. Oxidation and transamination of the 3"-position of UDP-*N*-acetylglucosamine by enzymes from *Acidithiobacillus ferrooxidans*: Role in the formation of lipid A molecules with four amide-linked acyl chains. *J. Biol. Chem.* **2004**, *279*, 25400-25410.
53. Zhao, G.; Winkler, M.E. A novel α-ketoglutarate reductase activity of the *serA*-encoded 3-phosphoglycerate dehydrogenase of *Escherichia coli* K-12 and its possible implications for human 2-hydroxyglutaric adicuria. *J. Bacteriol.* **1996**, *178*, 232-239.
54. Thoden, J.B.; Holden, H.M. Structural and functional studies of WlbA: a dehydrogenase involved in the biosynthesis of 2,3-diacetamido-2,3-dideoxy-D-mannuronic acid. *Biochemistry* **2010**, *49*, 7939-7948.
55. Nurizzo, D.; Halbig, D.; Sprenger, G.A.; Baker, E.N. Crystal structures of the precursor form of glucose-fructose oxidoreductase from *Zymomonas mobilis* and its complexes with bound ligands. *Biochemistry* **2001**, *40*, 13857-13867.
56. Wenzel, C.Q.; Daniels, C.; Keates, R.A.B.; Brewer, D.; Lam, J.S. Evidence that WbpD is an *N*-acetyltransferase belonging to the hexapeptide acyltransferase superfamily and an important protein for O-antigen biosynthesis in *Pseudomonas aeruginosa* PAO1. *Mol. Microbiol.* **2005**, *57*, 1288-1303.
57. Raetz, C.R.H.; Roderick, S.L. A left-handed parallel β-helix in the structure of UDP-*N*-acetylglucosamine acyltransferase. *Science* **1995**, *270*, 997-1000.
58. Vuorio, R.; Härkönen, T.; Tolvanen, M.; Vaara, M. The novel hexapeptide motif found in the acyltransferases LpxA and LpxD of lipid A biosynthesis is conserved in various bacteria. *FEBS Lett.* **1994**, *337*, 289-292.
59. Olsen, L.R.; Vetting, M.W.; Roderick, S.L. Structure of the *E. coli* bifunctional GlmU acetyltransferase active site with substrates and products. *Protein Sci.* **2007**, *16*, 1230-1235.
60. Olivier, N.B.; Imperiali, B. Crystal structure and catalytic mechanism of PglD from *Campylobacter jejuni*. *J. Biol. Chem.* **2008**, *283*, 27937-27946.

61. Schoenhofen, I.C.; McNally, D.J.; Vinogradov, E.; Whitfield, D.; Young, N.M.; Dick, S.; Wakarchuk, W.W.; Brisson, J.-R.; Logan, S.M. Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: Enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *J. Biol. Chem.* **2006**, *281*, 723-732.
62. Olivier, N.B.; Chen, M.M.; Behr, J.R.; Imperiali, B. In vitro biosynthesis of UDP-*N*,*N*'-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry* **2006**, *45*, 13659-13669.
63. Thoden, J.B.; Holden, H.M. Molecular structure of WlbB, a bacterial *N*-Acetyltransferase involved in the biosynthesis of 2,3-diacetamido-2,3-dideoxy-D-mannuronic acid. *Biochemistry* **2010**, *49*, 4644-4653.
64. Westman, E.L.; McNally, D.J.; Rejzek, M.; Miller, W.L.; Kannathasan, V.S.; Preston, A.; Maskell, D.J.; Field, R.A.; Brisson, J.-R.; Lam, J.S. Identification and biochemical characterization of two novel UDP-2,3-diacetamido-2,3-dideoxy-α-D-glucuronic acid 2-epimerases from respiratory pathogens. *Biochem. J.* **2007**, *405*, 123-130.

Chapter 4 Structural Analysis of WbpE from *Pseudomonas aeruginosa* PAO1: A Nucleotide Sugar Aminotransferase Involved in O-Antigen Assembly

A significant portion of the work described in this chapter has been published in the following:

Larkin, A.; Olivier, N.B.; Imperiali, B. Structural Analysis of WbpE from *Pseudomonas aeruginosa* PAO1: A Nucleotide Sugar Aminotransferase Involved in O-Antigen Assembly. *Biochemistry* **2010**, *49*, 7227-7237.

Introduction

Pseudomonas aeruginosa is a Gram-negative pathogen responsible for severe infection in immunocompromised individuals. In recent years, it has emerged as a major source of hospital-acquired infections, such as bacteremia, pneumonia and febrile neutropenia, and leads to inflammation and pulmonary failure in cystic fibrosis patients.[1-5] While there are several standard antibiotics currently used to combat *P. aeruginosa*, the rapidly increasing occurrence of multidrug resistant strains has complicated treatment and highlighted the need for alternative drug design strategies.[6-8] One of the most promising targets for improved treatment is the bacterial lipopolysaccharide (LPS), found in the exterior leaflet of the outer membrane of the organism. The LPS of *P. aeruginosa* contains the B-band O-antigen unit, a critical virulence factor that has been shown to play a key role in host colonization and evasion of immune defenses.[9-11] Genetic mutants of *P. aeruginosa* deficient in B-band O-antigen have exhibited greatly increased sensitivity to phagocytosis and serum-mediated killing.[12]

In *P. aeurginosa* PAO1 (serotype O5), the B-band O-antigen is composed of repeating units of a trisaccharide containing 2-acetamido-3-acetamidino-2,3-dideoxy-β-D-mannuronic acid (ManNAc(3NAm)A), 2,3-diacetamido-2,3-β-D-mannuronic acid (ManNAc(3NAc)A), and *N*-acetyl-α-D-fucosamine (Fuc2NAc) (Figure 3-1).[13] As outlined in Chapter 3, recent studies have shown that the central saccharide unit, ManNAc(3NAc)A, is synthesized by five enzymes in the Wbp pathway (WbpA, WbpB, WbpE, WbpD, and WbpI) starting from the common precursor, UDP-GlcNAc (Figure 4-1).[14] In the first step, WbpA oxidizes the C6" hydroxyl group to afford UDP-GlcNAcA. This is followed by the coupled action of the dehydrogenase WbpB and aminotransferase WbpE to yield UDP-GlcNAc($3NH_2$)A through a unique NAD^+ recycling pathway, in which the α-ketoglutarate product of WbpE is utilized by WbpB to

regenerate NAD^+.[15] Finally, the acetyltransferase WbpD and epimerase WbpI provide the final UDP-ManNAc(3NAc)A donor, which is then transferred to an undecaprenyldiphosphate carrier. After completion, the trisaccharide is flipped into the bacterial periplasm by the Wzx translocase and then ligated to the growing O-antigen by the coordinated action of the Wzy polymerase and the chain length regulator Wzz.[16]

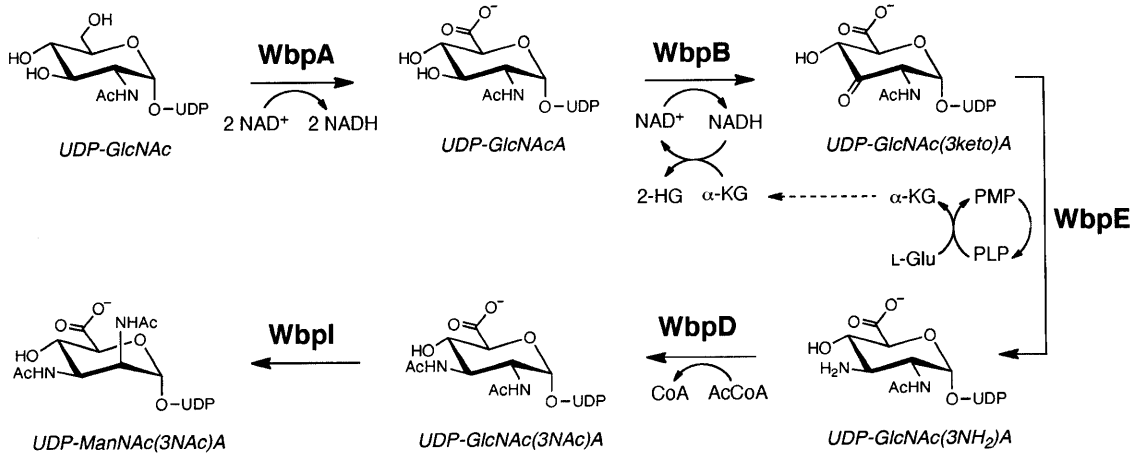


Figure 4-1: Biosynthetic pathway of UDP-ManNAc(3NAc)A in *P. aeruginosa* PAO1.

WbpE is a pyridoxal 5'-phosphate (PLP)-dependent aminotransferase responsible for the conversion of UDP-GlcNAc(3keto)A and L-glutamate to UDP-GlcNAc($3NH_2$)A and α-ketoglutarate (α-KG), respectively. It is a member of the broad class of Fold Type I aspartate-aminotransferase (AAT) enzymes, which harness the powerful electron-sink properties of PLP to carry out a wide variety of transformations, including transaminations, eliminations, decarboxylations, and epimerizations.[17,18] The general mechanism of this class of enzymes has been worked out in great detail, and is divided into two discrete half reactions that cycle between the PMP and PLP forms of the cofactor (Figure 4-2). The reaction begins with imine

formation between the ketone-bearing sugar and PMP, followed by abstraction of a C4' proton of PMP by the catalytic lysine residue of the protein. The resulting quinoid intermediate then accepts a proton from the lysine at the C3" carbon of the pyranose, which serves to set the stereochemistry at this position and results in the external aldimine intermediate. Release of the nucleotide sugar amine product is facilitated by transimination to afford the internal aldimine. In the second half of the catalytic cycle, the reaction steps are repeated in the reverse order resulting in regeneration of PMP through conversion of L-glutamate to α-ketoglutarate.

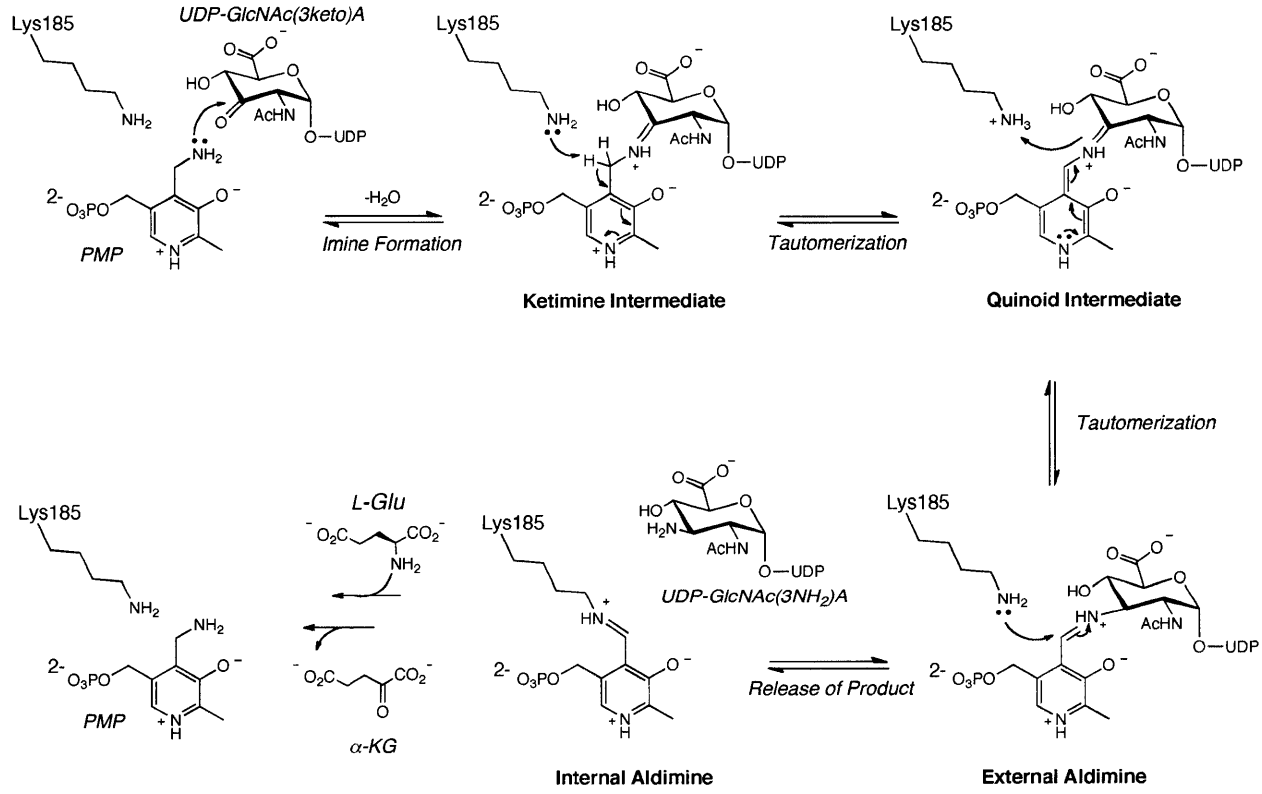


Figure 4-2: Proposed reaction mechanism of WbpE.

The first crystal structure of a Fold Type I aminotransferase reported was that of aspartate aminotransferase (AAT), an enzyme that catalyzes the conversion of L-aspartate and α-keotglutarate to oxaloacetate and L-glutamate.[19] Found in high levels in the mitochondria of hepatic cells, the upregulation of this enzyme is often correlated with liver inflammation and is used as a standard clinical biomarker of disease.[20,21] In 1980, AAT was crystallized by the Jansonius group as a homodimer in complex with its PLP cofactor (Figure 4-3A);[22] since the publication of this landmark paper, over forty crystal structures of Fold Type I aminotransferases of have been elucidated and deposited in the Protein Data Bank (PDB). Comparison of these structures reveals several defining features that enable PLP-binding and catalysis of the wide variety of reactions carried out by these enzymes. For example, a large number of these enzymes are compact oligomers, often homodimers, in which the PLP-binding sites are open to face one another. In addition, nearly all known Fold Type I aminotransferases contain a unique central β-sheet comprised of seven β-strands, with a topology in which the seventh strand runs antiparallel to the remaining six (Figure 4-3B).

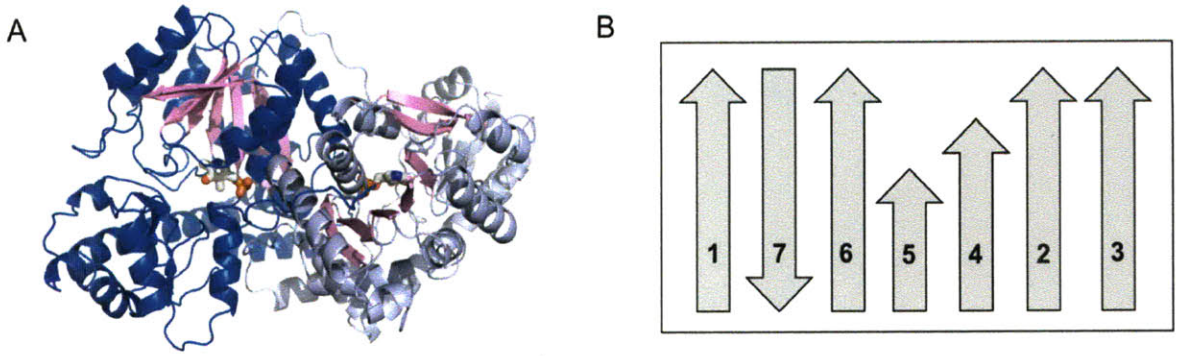


Figure 4-3: (A) Crystal structure of aspartate aminotransferase, with the PLP cofactor shown as sticks and the central β-sheet colored pink (PDB code: 1ASM); (B) Organization of β-strands within the canonical β-sheet found in the Fold Type I family of aspartate aminotransferases.

In recent years, the crystal structures of several Fold Type I enzymes that catalyze the transamination of nucleotide sugars in bacteria have been described.[23-28] Despite the similarities observed in the overall protein scaffold of these structures, several questions regarding substrate binding and specificity still remain. In particular, it is unclear how these enzymes achieve their exquisite specificity for both the nucleotide sugar and donor amino acid substrates, as the residues lining the active sites of these proteins are quite similar. Several hypotheses have been put forth to explain this finding, including the existence of a conformational change in the enzyme structure upon substrate binding, as well as an increased affinity for certain substrates due to induced fit.[18] However, comparison of these structures in the apo state with those in complex with the nucleotide sugar substrate do not show significant evidence of a conformational change and are not suggestive of an induced fit, thus casting doubt on these hypotheses.[29] Therefore, further work is required to gain a better picture of the determinants that affect binding and specificity of this class of enzymes.

In this chapter, we present the crystal structure of the aminotransferase WbpE from *P. aeruginosa* PAO1 in complex with three different ligands: PMP, PLP, and the external aldimine of PLP with the reaction product, UDP-GlcNAc(3NH_2)A. This work presents the first structure of an aminotransferase that binds a nucleotide sugar modified at the C2", C3", and C6" positions, and thus provides a glimpse into the means by which such a highly functionalized nucleotide sugar is accommodated in the enzyme active site. In addition, it was confirmed that the dimeric structure of WbpE observed in the crystal structure is also present in solution using sedimentation velocity analytical ultracentrifugation. In order to gain a broader understanding of nucleotide sugar specificity across this class of aminotransferases, the WbpE structure in complex with the external aldimine is compared to other nucleotide sugar-bound

aminotransferase structures. Efforts toward obtaining a crystal structure of a putative complex of both the dehydrogenase WbpB and WbpE are also summarized, with the goal of gaining structural insight into the interactions between these two enzymes that facilitate the novel NAD^+ recycling behavior outlined in Chapter 3. We envision that this work will add to the growing body of knowledge about nucleotide sugar aminotransferases and may also present a novel target for the development of antimicrobial treatments to combat *P. aeruginosa.*

Results and Discussion

Overexpression, Purification, and Crystallization of Native WbpE

The *wbpE* gene was amplified from *P. aeruginosa* PAO1-LAC genomic DNA and inserted into a modified form of the pET32a vector using standard molecular biology techniques. The resulting protein contained an N-terminal His_6-tag followed by a tobacco etch virus (TEV) protease cleavage site and was overexpressed in very high yield, with over 85 mg of purified WbpE obtained from 1 L cell culture. After purification by Ni-NTA resin, the N-terminal His_6-tag was removed by incubation with TEV protease at 4 °C. The removal of the tag was confirmed by Western blot analysis with an Anti-His_4 antibody, and the protein then subjected to size-exclusion chromatography to achieve highly pure, monodispersed protein. The final product was analyzed by SDS-PAGE to confirm purity prior to setting up crystal trays.

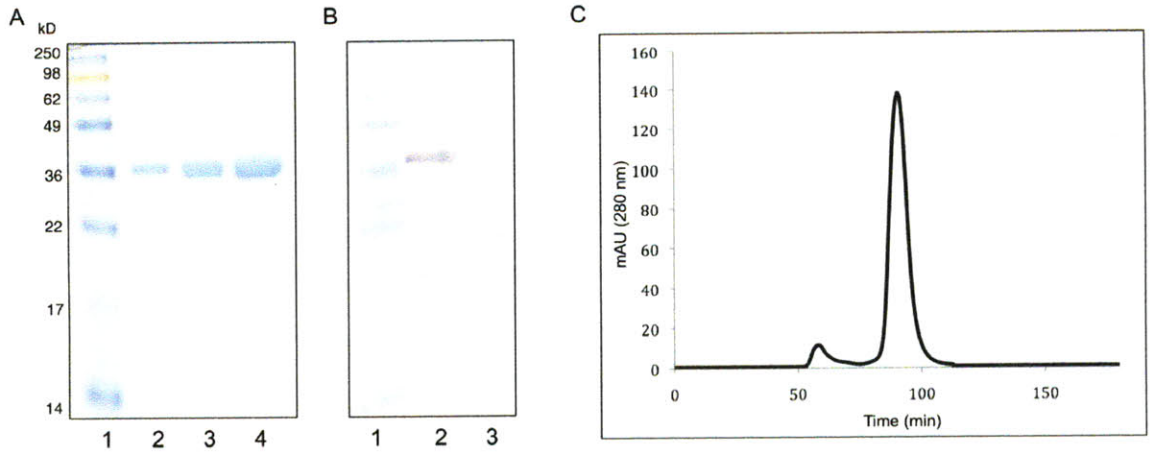


Figure 4-4: A) 12% SDS-PAGE of purified WbpE, with increasing amounts of protein (2, 4, and 6 μL) loaded; (B) Anti-His_4 Western blot, depicting WbpE before (2) and after (3) TEV cleavage to remove the His_6-tag. The band in Lane 3 corresponds to His_6-tagged TEV protease present in the reaction; (C) Size-exlcusion chromatogram representing WbpE after TEV cleavage. Fractions eluting between 85 and 100 minutes represent monodispersed protein and were utilized for crystallography.

Crystallization conditions for WbpE were identified by initially surveying several commercially available sparse matrix screens. In general, samples of WbpE were prepared at concentrations of 10, 15, and 20 mg/mL and were set up at a 1:1 ratio of protein to crystallization condition in a 96-well sitting drop format at 25 °C. Crystal trays were monitored daily, and after two weeks, the formation of small, square crystals was observed for several conditions. Optimization of these conditions led to large, oval-shaped crystals that diffracted to 1.4 Å. These crystals were transferred to a cryoprotectant solution containing 20% glycerol prior to X-ray data collection; representative crystals are depicted in Figure 4-5.

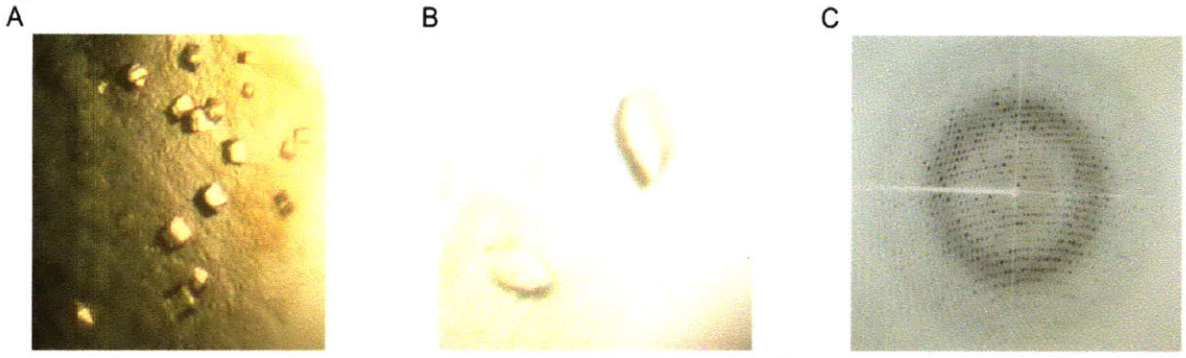


Figure 4-5: Crystallization of native WbpE. (A) Initial crystals of WbpE obtained from a sparse matrix screen, which were optimized to obtain diffraction quality crystals shown in (B); (C) A sample diffraction pattern of native WbpE (1.6 Å).

Obtaining Phasing Information to Solve the WbpE Structure

All attempts to solve the crystal structure of native WbpE using molecular replacement with homologous nucletotide sugar aminotransferases, such as DesV (*Streptomyces venezuelae,* sequence identity = 41%) and PseC (*Helicobacter pylori*, sequence identity = 27%),[24,26] as the initial search models were unsuccessful. Therefore, an alternative method for obtaining phasing information was sought. As a first attempt, native WbpE crystals were soaked in various solutions of heavy atoms from a commercially available screen with the goal of solving the structure using the method of multiple isomorphous replacement (MIR).[30] MIR is a widely used means of gaining phasing information through the use of heavy atom derivatives, in which a heavy atom such as gold, mercury, or lead replaces another atom in a macromolecule to produce an isomorphous derivative. Traditionally, these derivatives are generated by either soaking a native crystal in a solution containing heavy atoms or co-crystallizing the protein with the solution of choice. The basic principle behind this method involves combining structure factors from the native data set with those obtained from heavy atom data sets and identifying the differences with the aid of Patterson maps.[31] Despite an extensive screen of soaking conditions,

in which the concentration of heavy atom, time, and cryoprotectant concentration were varied, a solution of WbpE was not obtained (Table 4-1). In all cases, either the WbpE crystals were either unable to withstand the soaking conditions and rapidly disintegrated, or the heavy atom signal was not observed upon data set analysis with the software programs SOLVE or SHELXD.[32,33]

Table 4-1: Summary of heavy atom soaking trials of WbpE crystals for MIR

Heavy Atom Soak[a]	**Diffraction Quality**	**Outcome**[b]
$SrCl_2$	very poor	unable to use data set
Na_2SeO_4	good	only slight incorporation of Se
$K_2Pt(CN)_4$	good	no evidence of Pt by FES
K_2PtCl_6	good	no evidence of Pt by FES
$KAuCl_4$	no diffraction	no evidence of Au by FES
$PbOAc_2$	good	only slight incorporation of Pb
$HgOAc_2$	no diffraction	no evidence of Hg by FES
$HgEtH_2PO_4$	no diffraction	no evidence of Hg by FES
$(NH_4)_2WS_4$	poor	no evidence of W by FES
TlCl	poor	no evidence of Tl by FES
K_2IrCl_6	poor	no evidence of Ir by FES
$HoCl_3$	good	evidence of Ho by FES

[a]Soaking conditions involved incubation of native WbpE crystals in solutions containing increasing concentration of the heavy atom. Crystals were initially soaked for 10 mins while monitoring crystal appearance; further soaks were carried out for incubation times up to 48 hours.
[b]FES = fluorescence energy scan

As an alternative to MIR, phasing information was sought through the incorporation of selenomethionine (SeMet) into the protein sequence. Pioneered by Wayne Hendrickson in the early 1990s,[34] this method allows for the replacement of native methionine residues with SeMet,

thus allowing for structure determination by multiple anomalous dispersion (MAD). It is generally accepted that in most cases, SeMet is able to replace methioinine without causing a noticeable disturbance in the overall protein structure.[35] In addition, knowledge of exact selenium atom locations in the protein sequence serves as a significant aid in model building.

Over the past two decades, several procedures have been developed in order to generate SeMet-derivatized protein;[36] three of these methods were attempted in order to introduce SeMet into the WbpE sequence. The first method was initially described by Guerrero et al, and involved the inoculation and growth of *E. coli* cells transformed with the WbpE-pET32a plasmid following standard procedures in Luria-Bertani (LB) broth.[37] After reaching the appropriate culture density, the cells were transferred to a minimal medium in which L-SeMet was provided. WbpE was purified from these cells after induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by overnight expression, and efficiency of SeMet incorporation was analyzed by MALDI mass spectrometry. Although the level of SeMet labeling was near 100%, the overall protein yield was poor (< 3 mg/L) and thus alternate means of protein expression were explored.

A second attempt to achieve expression of a SeMet-derivative of WbpE involved the use of a methionine auxotroph. The B834 (DE3) *E. coli* cell line, which contains a lesion in the *metA* gene and thus is unable to synthesize methionine, was obtained from a commercial source and transformed with the WbpE-pET32a plasmid. The cells were then cultured in LB media in the presence of SeMet following standard protocols. As in the case with the previous method, MALDI mass spectrometry analysis showed nearly full incorporation of SeMet, but the overall protein yield was prohibitively low (1 mg/L) and thus prevented crystallization efforts.

A third and ultimately successful method of incorporating SeMet into WbpE involved the process of metabolic inhibition. Previous work has shown that high levels of certain amino acids, such as leucine, isoleucine, and lysine, serve to shut down the native methionine biosynthesis pathway in bacterial cells through the inhibition of aspartokinases, thus rendering the cells more likely to take up SeMet from exogenous sources.[36] In the last two decades, several protein structures were solved using this technique, including FKBP-12[38] and UDP-acetylenylpyruvylglucosamine reductase.[39] The basic procedure involves the inoculation and growth of standard BL21(DE3) RIL *E. coli* cells transformed with the WbpE-pET32a plasmid in LB media until the appropriate culture density is reached. At this point, a combination of seven amino acids (L-isoleucine, L-leucine, L-valine, L-threonine, L-phenylalanine, and L-lysine, and L-SeMet) were added to the culture flask to initiate inhibition of methionine biosynthesis; after 20 mins, IPTG was introduced to the culture to induce expression of the SeMet-labeled protein. Using this procedure, SeMet-WbpE was overexpressed in very high yield, with over 45 mg of protease-treated, purified protein obtained from a 500 mL cell culture. Analysis of the labeled WbpE by MALDI mass spectrometry indicated full occupancy of SeMet.

Crystallization of SeMet-WbpE was performed in the same manner as that described for native WbpE, in which a purified protein sample was surveyed with commercially available sparse matrix screens in a sitting drop 96-well plate format at 25 °C. Optimization of initial crystallization conditions involved varying the pH and precipitant concentration as well as the use of an additive screen; it was found that the addition of 10 mM $SrCl_2$ into the crystallization drop helped to generate large, oval-shaped crystals similar to those seen for native WbpE. Diffraction quality crystals were then obtained in a well solution comprised of 0.2 M NH_4OAc, 0.1 M Bis-Tris (pH 5.5), 10 mM $SrCl_2$, and 25% PEG 3350 and diffracted to 1.8 Å.

Data Collection and Structure Determination

Diffraction data were collected at 110 K on beamline X6A at the National Synchrotron Light Source at Brookhaven National Laboratory (Upton, NY). For the SeMet derivative of WbpE, data were accumulated at the selenium peak (12667 eV), absorption edge (12660 eV) and remote energies (12867 eV). Data sets were both indexed and scaled using HKL2000, and the scaled intensities were converted to structure factors using the program TRUNCATE in the Collaborative Computational Project Number 4 (CCP4) suite of programs.[40,41] Data collection parameters are summarized in Table 4-2.

The structure of the WbpE-SeMet derivative was solved using the method of multiwavelength anomalous diffraction (MAD) as previously described.[34] Using data collected at the selenium peak and truncated to 2.5 Å, the program SOLVE was employed to locate the heavy atom sites and generate experimental phases.[32] A total of four out of five selenium atoms were located for each protein subunit. The initial model was built using the prime-and-switch phasing feature of RESOLVE to minimize model bias and contained the basic framework of the model,[42,43] except for sizeable gaps between residues 116-142 and 196-233. The PMP-bound structure was then solved by molecular replacement using Phaser with the SeMet derivative as the initial search model.[44] Further model building and refinement were carried out using Coot and Refmac.[45,46] Five percent of the data were used to calculate the R_{free} values for cross-validation of the refinement process.[47] The structures of the PLP and external aldimine complexes were subsequently solved by molecular replacement in the same manner, using the completed PMP-bound structure with the cofactor removed as the input model. Water molecules were added using both ARP/warp and Coot, and ligands were modeled into each structure after

the R_{free} value was below 30%.[48] All refined structures were validated using PROCHECK, SFCHECK, and MolProbity,[49-51] and the final refinement statistics are presented in Table 4-2.

Table 4-2: Data Collection and Refinement Statistics

	PMP	PLP	External Aldimine	SeMet[a]
Data Collection				
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell dimensions (*a*, *b*, *c*) (Å)	78.28, 149.29, 54.90	77.79, 148.72, 53.22	78.29, 149.67, 55.12	77.81, 149.03, 53.29
Resolution (Å)	50.0-1.95	20.0-1.50	20.0-1.83	50.0-1.93
Observed reflections	45,308	94,182	46,379	46,078
R_{merge} (%)[b,c]	7.3(42.0)	9.1(66.2)	7.1(49.2)	8.2(41.9)
$I/\sigma I$[c]	53.4(9.2)	35.8(3.0)	36.3(4.1)	36.7(5.8)
Completeness (%)[c]	99.8	99.7	98.6	100
Redundancy[c]	14.6(14.5)	7.9(7.0)	7.3(6.7)	7.5 (7.2)
Refinement				
Resolution (Å)	35-1.95	20-1.50	20-1.90	
R_{work}/R_{free} (%)[a,e]	19.8/24.6	17.9/22.2	18.7/24.5	
Total Number of Atoms	5837	5880	6027	
Protein	5428	5471	5401	
Water	377	409	516	
Ligands	32	32	110	
B-factors (Å^2)				
Overall	28.2	20.5	19.8	
Protein	28.3	19.1	19.3	
Water	29.4	22.7	27.4	
Ligand	15.0	16.9	34.8	
Ramachandram plot (%)[f]	96.8/2.8/0.4	97.1/2.7/0.3	96.0/3.8/0.3	
r.m.s. deviations				
Bond lengths (Å)	0.010	0.011	0.016	
Bond angles (°)	1.245	1.250	1.350	
PDB code	3NU7	3NU8	3NUB	

[a] Data reported for the SeMet derivative refer to those collected at the selenium peak wavelength.
[b] The number in parathenses represents the highest resolution bin; 1.98-1.94, 1.55-1.5, 1.86-1.83 and 2.01-1.93 Å, for the data sets of PMP, PLP, external aldimine and the SeMet derivative, respectively.
[c] $R_{merge} = \Sigma||I-\langle I\rangle|/\Sigma I$, where I is the intensity of a reflection, and $\langle I\rangle$ is the mean intensity of group of equivalent reflections.
[d] $R_{work} = \Sigma_h||F(h)_{obs}|-|F(h)_{calc}||/\Sigma_h|F(h)_{obs}|$
[e] R_{free} was calculated for 5% of the reflections randomly excluded from the refinement.
[f] Ramachandran plot statistics are given as core/allowed/generously allowed and are for both chains.

Overall Architecture of WbpE

WbpE crystallized in the orthorhombic space group $P2_12_12$, with two molecules in the asymmetric unit and approximate unit cell dimensions of 75 Å x 150 Å x 50 Å (Figure 4-6). As predicted, the overall scaffold of WbpE is similar to that of other members in the Fold Type 1

aspartate aminotransferase family.[19,22] As in the case of AAT, each monomer of WbpE can be divided into two major domains: a large N-terminal PLP-binding region and a smaller C-terminal domain, shown in purple and green in Figure 4-6C. The N-terminal portion (residues Ile14-Glu243) is composed of a total of eleven β-strands, seven of which make up a central β-sheet that exhibits the canonical strand order β-1, β-9, β-8, β-5, β-4, β-2, β-3, where β-9 is antiparallel to the others. This β-sheet is bordered on the top and bottom faces by seven α-helices. After strand β-9, a short stretch of residues (Gln209-Arg229) extends out from the monomer structure to form a large domain-swapped β-hairpin (β-10 and β-11) that interacts with the active site of the other monomer. Another smaller β-hairpin (residues Ala163-Ser170, β-6 and β-7) protrudes from the top face of the central β-sheet and interacts with the C-terminal region in the same subunit. The C-terminal portion of the protein (residues Met1-Arg13, Ile244-Asn359) is made up of two short β-strands and six α-helices that are oriented away from each other in a V-shape. The entire C-terminal region of the protein is closely aligned to the N-terminal domain on the top face of the monomer, yet opens to reveal a deep channel on the opposite face of the subunit, allowing solvent access to the enzyme active site (Figure 4-6B). In addition, this domain of the protein contains a non-prolyl *cis* amide bond between residues His308 and Tyr309.

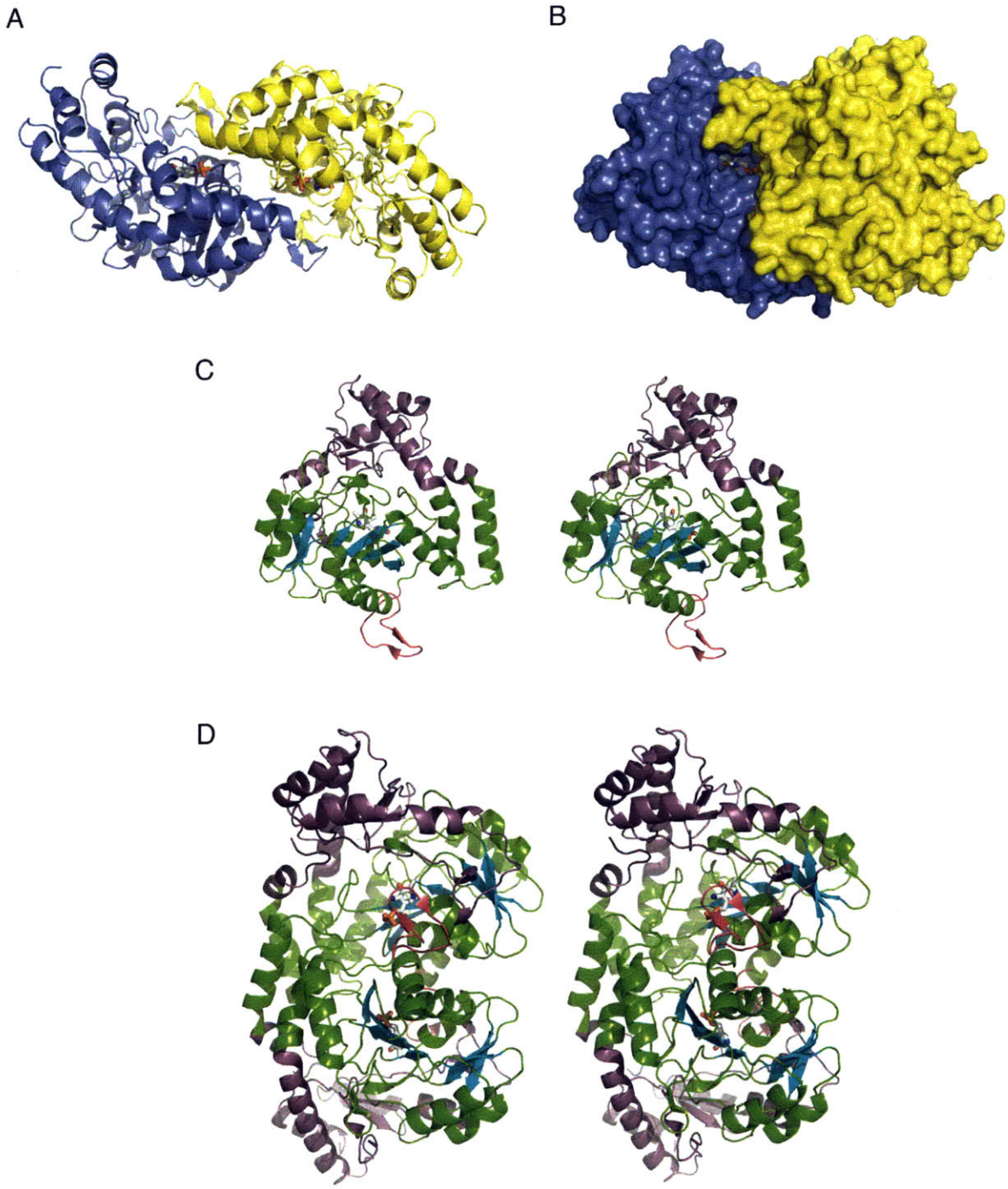


Figure 4-6: Crystal structure of the WbpE-PMP complex. (A) Structure of the homodimer, with each subunit colored individually and the PMP cofactor shown as sticks. (B) Space-filling model, where the PMP cofactor can be observed at the end of a deep cavity. (C) Stereodiagram of a single subunit, with the N- and C-terminal domains colored green and purple, respectively. The β-strands comprising the central β-sheet are highlighted in cyan, and the domain-swapped β-hairpin is colored magenta. (D) Stereodiagram of the homodimer, with the various domain features colored as described for panel C.

The interface between the subunits of the dimer is extensive and comprises an area of approximately 7500 $Å^2$, which represents nearly 25% of the total surface area of a single subunit. The two active sites of the dimer are open to one another and are spaced roughly 30 Å apart. The most striking interaction between the two monomers is the presence of the domain-swapped β-hairpin that crosses over the dimer interface and lines the entrance to the active site tunnel of the neighboring subunit. Other notable interactions between the monomers include the stacking of two homologous antiparallel α-helices from each monomer, both comprising residues Ile14-His28, as well as the loop residues Cys189-Gly193 from one subunit that nestle beneath residues Gly29-Leu33 from the other (Figure 4-6).

Sedimentation velocity analytical ultracentrifugation studies were undertaken in order to establish whether the dimerization of WbpE observed in the crystal structure correlates with the oligomeric state present in solution. The sedimentation coefficient was found to be 4.72 Svedberg units and the molecular mass was calculated as 79.3 kD using a single species model fit. As each monomer of WbpE has a theoretical molecular weight of 38.9 kD, the data suggest that WbpE associates as a dimer in solution (Figure 4-7). This finding is in keeping with other members of the Fold Type 1 aminotransferase family, which have been found to exist primarily as homodimers or other higher-order oligomers in mutiples of two.[18] In addition, this result suggests that the dimeric structure of WbpE observed in the crystal structure correlates with the native structure of the enzyme in solution.

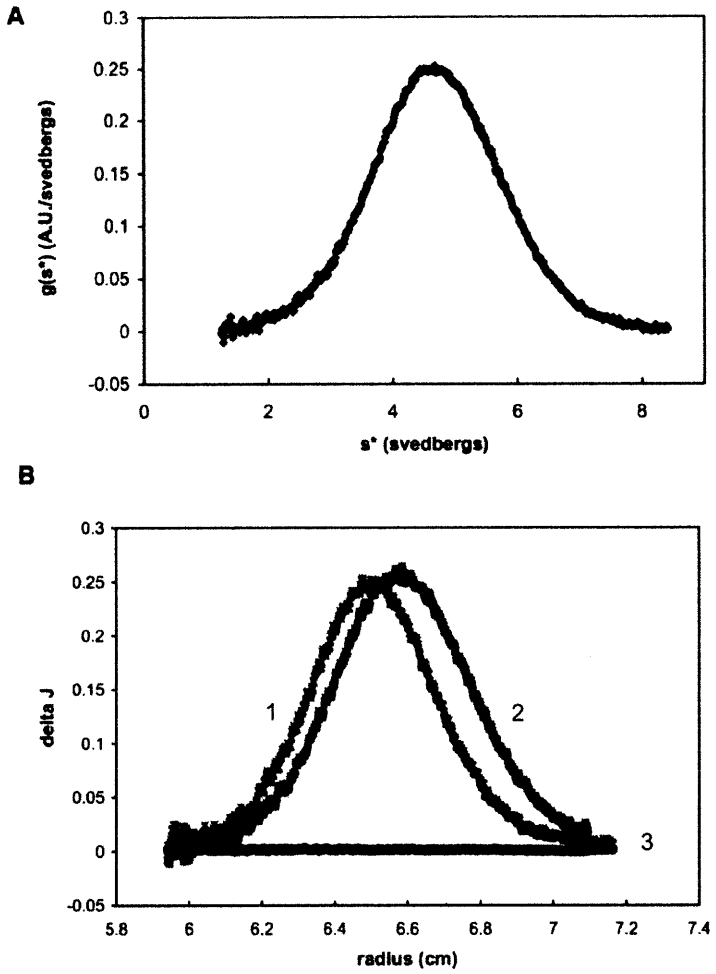


Figure 4-7: Sedimentation velocity analytical ultracentrifugation data. (A) Time derivative (g(s*)) plot of WbpE depicting a sedimentation coefficient of 4.72 svedbergs. (B) Time difference residual for WbpE. The raw data (1), corresponding model (2), and residuals of the fit (3) for two pairs of scans taken over the course of the experiment.

Cofactor Binding Site

During structure refinement, both the SeMet derivative and native form of WbpE were discovered to contain the cofactor bound to the enzyme active site despite the fact that it was not exogenously introduced over the course of protein expression or purification. The electron density of the cofactor was unambiguous and thus modeled into the native structure at full occupancy (Figure 4-8A). It was determined that the cofactor was present in the PMP form due

to the retracted position of the catalytic lysine residue (Lys185) as well as the lack of electron density between the C4' carbon of the cofactor and the lysine nitrogen atom, indicating that there was no covalent bond present. In addition, the UV spectrum of purified WbpE possessed the spectral characteristics of PMP, namely the presence of an absorbance maximum at 340 nm compared with that at 420 nm indicative of PLP. In order to obtain the PLP-bound form of the enzyme (internal aldimine), PLP was incubated with WbpE for 1 hr prior to setting up crystal trays. In contrast to the PMP-bound crystals, the PLP-bound structure showed an extended conformation of Lys185 and clear electron density between the side chain and the cofactor, serving as evidence that the cofactor was held in place by a covalent bond (Figure 4-8B).

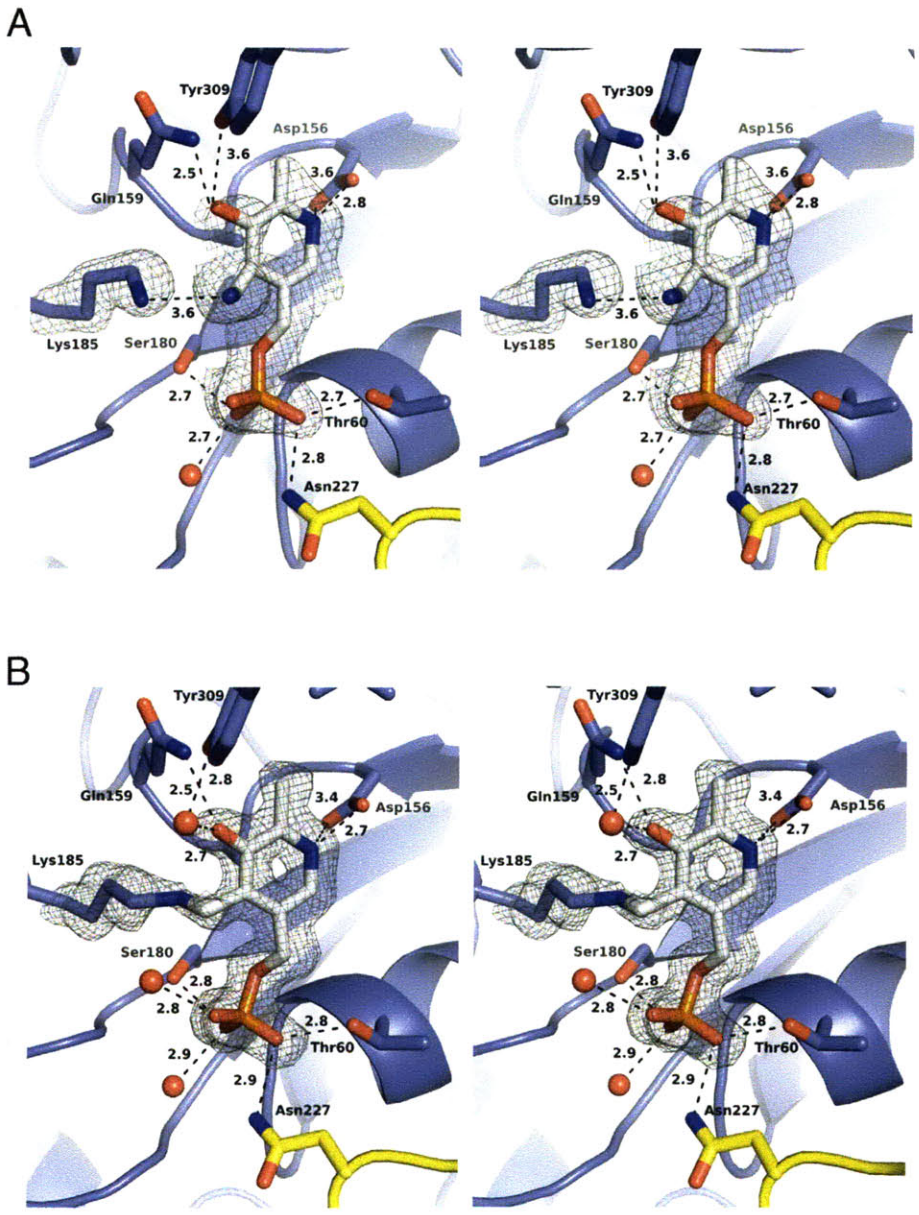


Figure 4-8: Stereodiagram of the cofactor binding site. The $2(F_o\text{-}F_c)$ electron density maps are contoured at 3σ for both the PMP-bound (A) and PLP-bound (B) structures and were calculated with atoms of the cofactor and Lys185 side chain omitted. Amino acid residues corresponding to subunits 1 and 2 are colored blue and yellow, respectively, and water molecules are depicted as red spheres. Possible hydrogen bonding interactions are shown by dashed lines All nearby residues within 3.6 Å of the cofactor are indicated except for those that line the top face of the cofactor-binding site; these residues were removed from the figure for the sake of clarity but are outlined in the text.

The cofactor-binding site resides within a deep cleft in the interior of each monomer at the far end of the joined active sites. The cofactor is oriented such that the pyridinium ring is facing inward, resting above the central β-sheet of each monomer unit, while the phosphate moiety points toward the entrance of the cleft and is located 20 Å from the phosphate group in the other subunit. The pyridinium ring is flanked on the top face by Tyr85, with which it is interacting in a π-π stacking fashion, and the bottom face by Ala158. The C5' hydroxyl group of the cofactor is within hydrogen-bonding distance of Gln159 and Tyr309, while the C6' methyl group is in close proximity (3.3 Å) to Val130. The N1 nitrogen of the pyridinium ring is held in place by a salt bridge with the carboxylate of Asp156 that serves to enhance the electron-sink properties of the cofactor. This aspartate residue is conserved in all Fold Type I aminotransferases and is crucial for maintaining the cofactor in its protonated state.[17] In the PMP-bound form of the enzyme, the C4' amine does not appear to interact with any specific residues, as the closest atom is the oxygen of a water molecule more than 4 Å away. Examination of a space-filling model of the PMP-bound form of WbpE shows that the C4' amine lies at the bottom of the deep solvent tunnel leading to the outside of the active site, making it accessible for reaction with incoming substrate (Figure 4-6B). The phosphate group of the cofactor protrudes into the entrance to the active site in the opposite direction and is held in place by several hydrogen-bonding and charge stabilizing interactions, including those with Ser180, Thr60, Gly59, Asn227 of the β-hairpin from the other subunit and three well-ordered water molecules.

An analysis of the residues surrounding the cofactor in both the PMP and PLP-bound structures indicate that there is very little conformational change in the overall architecture of WbpE upon formation of the internal aldimine (Figure 4-8); the root mean square deviation of

these two structures is 0.32 Å. Differences include the slight upward shift of the pyridinium ring into the active site due to the presence of the imine bond as well as small changes in the orientation of the Tyr85 side chain. In addition, the domain-swapped β-hairpin stretches closer toward the active site in the PLP-bound structure; slight changes can be seen in the side chain orientations of Gln215 and Arg212 that reflect this movement.

Nucleotide Sugar Binding Site

Crystals of WbpE complexed with the external aldimine were obtained by incubating the enzyme with both PLP and the UDP-GlcNAc(3NH_2)A product on ice for 1 hr prior to crystallization. The presence of the external aldimine in the enzyme active site implies that the reaction proceeded in reverse upon exposure to the aminated product; this intermediate was previously observed in other nucleotide sugar-bound aminotransferase structures as well.[24,25,28] The external aldimine is bound in both subunits of the dimer, and electron density between the cofactor pyridinium C4' carbon and the C3" position of the hexose ring indicates the existence of a covalent bond (Figure 4-9A). As in the case of the WbpE complexes with PMP and PLP, the electron density for the cofactor portion of the ligand is very strong; however, the density for the nucleotide sugar component is weaker, with an average temperature factor $\langle B \rangle$ of 42.8 $Å^2$ compared with that for the cofactor (26.8 $Å^2$) and nearby protein atoms (21.6 $Å^2$). The weakest regions of electron density in the external aldimine structure are surrounding the C4" hexose carbon as well as the C2' and C3' carbon bond in the ribose moiety. This observation suggests that the nucleotide sugar region of the aldimine is not as firmly held in place as the cofactor and may experience some thermal motion within the crystal. For this reason, atoms comprising the nucleotide sugar were refined at 50% occupancy.

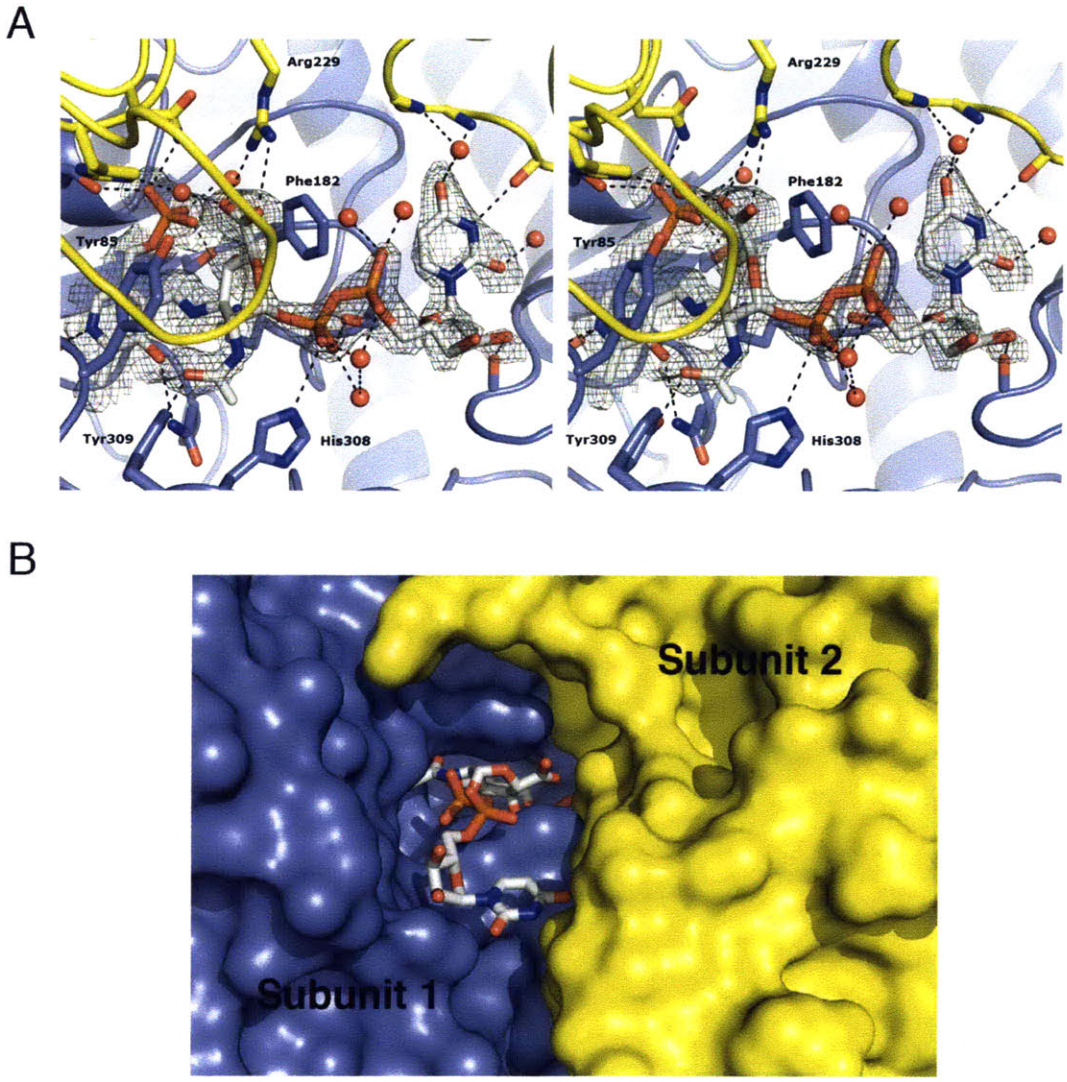


Figure 4-9: Stereodiagram of the external aldimine-bound structure. (A) External aldimine of PLP and UDP-GlcNAc(3NH_2)A in the active site of WbpE. The 2(F_o-F_c) electron density map is contoured at 1.5σ and was calculated with atoms of the external aldimine omitted. Residues within a 4 Å radius of the ligand are indicated, and water molecules are depicted as red spheres. (B) Space-filling model of the WbpE external aldimine complex, highlighting the manner in which the uridine moiety extends up through the deep cavity to the solvent. In both figures, the subunits of the dimer are colored blue and yellow.

The glucopyranose ring of the ligand is bound to the cofactor through the C3" carbon and, as depicted in Figure 4-9, many interactions take place between this highly functionalized sugar and the protein to accommodate it within the active site. The aromatic side chains of Tyr85 and Phe182 are stacked above and below the glucopyranose ring, presumably forming a

barrier to prevent solvent access throughout the course of catalysis. In addition, these residues may serve to stabilize the positively charged reaction intermediates through cation-π interactions. The C2" acetamido oxygen appears to participate in hydrogen bonds with both the side chain phenol of Tyr309 as well as the C5' hydroxyl group of the cofactor pyridinium ring, while the C4" hydroxyl is within hydrogen-bonding distance of the cofactor phosphoryl group. The C6" carboxylate is coordinated to a well-ordered water molecule that is hydrogen-bonded to both the phosphate group and the imidazole ring of His213, which is part of the domain-swapped β-hairpin from the neighboring protein subunit. In addition to this water molecule, the carboxylate is held in place by a salt bridge with the guanidinium moiety of Arg229, also from the β-hairpin domain.

The pyrophosphate of the nucleotide sugar extends away from the glucopyranose through the deep protein cavity that leads to the exterior of the protein, and is anchored by interactions with Ser184, His308 and at least two well-ordered waters. The uridine emerges from this channel and is fully exposed to the solvent, making van der Waals contacts with one α-helix from each of the dimer subunits and residues from both monomer units on the bottom face and several water molecules from above. The ribose 3'-OH group participates in hydrogen bonds with the N-terminal region of the protein, including the side chain of Glu3 and the backbone carbonyl of Ile5, while the uracil is in close proximity to numerous residues in the neighboring subunit, including Gly29, Tyr31 and Ile32. Comparison of the WbpE complexes with PMP and the external aldimine reveals that as in the case with the internal aldimine structure, no major conformational change is observed in the overall architecture of the protein upon substrate binding. Analysis of the overlaid cofactor and external aldimine-bound structures indicates only small differences between the two forms of the enzyme (r.m.s.d. = 0.33 Å), with

the most notable changes within the domain-swapped β-hairpin shifting slightly around the enzyme active site.

Perhaps one of the most interesting findings from this study is the coordination of an active site arginine, Arg229, with the C6" carboxylate of the external aldmine ligand. This arginine residue is commonly found in Fold Type I aminotransferases and has been shown to coordinate with the carboxylate of L-glutamate and α-KG in the first half of the catalytic cycle.[18] In the second half of the cycle, this arginine side chain is often found flipped out of the active site in order to make room for the second substrate, which often requires more space than L-glutamate or α-KG (Figure 4-10A). This back and forth movement has been called "arginine switching," and has been studied in other Fold Type I aminotransferases, such as tyrosine aminotransferase from *Paracoccus denitrificans*.[18,52] In the case of the nucleotide sugar aminotransferase ArnB from *Salmonella typhimurium*, a crystal structure of the protein in complex with α-KG and PLP shows the arginine positioned out of the active site, switched in the "off" position to make room for the nucleotide sugar substrate (Figure 4-10B).[23] Comparison of the WbpE active site with homologous protein structures implicates Arg229 as the arginine in question. However, unlike these other proteins, it appears that this Arg229 of WbpE does not possess a switching capability, for it is bound to the nucleotide sugar when it would be predicted to be oriented out of the active site. Presuming that α-KG is present in the same location in WbpE as it is in ArnB, it can then be inferred that Arg229 of WbpE interacts with both enzyme substrates (L-glutamate and UDP-GlcNAc(3keto)A), playing a critical role in both halves of the catalytic cycle. This alignment of α-KG and the UDP-GlcNAc(3keto)A substrates within the active site of WbpE complements previous work in our laboratory on this pathway, in which we postulated that the unique NAD^+ cofactor regeneration exhibited by the dehydrogenase WbpB

was facilitated by the overlap of α-KG and UDP-GlcNAc(3keto)A within the substrate-binding pocket (Chapter 3).[15]

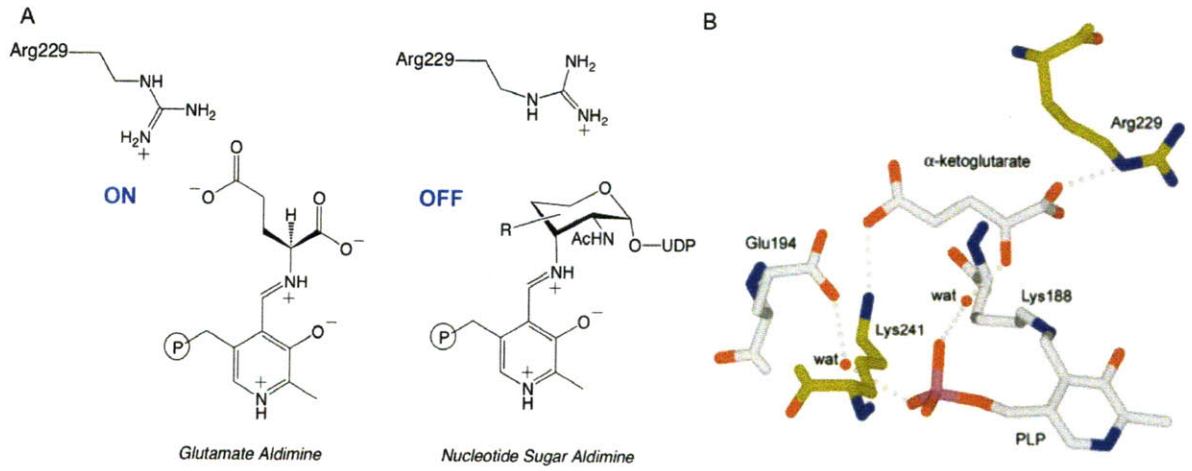


Figure 4-10: Examination of arginine-switching in Fold Type I aminotransferases. (A) Basic concept of switching behavior, wherein the arginine interacts with the L-glutamate substrate but moves out of position to accommodate the nucleotide sugar substrate. (B) Crystal structure of ArnB from *S. typhimurium* in complex with α-KG and PLP, showing the Arg229 residue oriented out of the active site in the "off" position after the first half of the catalytic cycle, primed for binding to the incoming nucleotide sugar substrate. Figure adopted from Noland et al (PDB code: 1MDX).[23]

Analysis of the Effect of Point Mutations on the Activity of WbpE

Comparison of the cofactor and nucleotide sugar-bound structures of WbpE led to the identification of eight key amino acid residues that appear to play a critical role in substrate binding. These residues are found in both subunits of the dimer and contact the cofactor and/or substrate either through hydrogen bonding (Thr 60, Gln159, Ser 180, Asn227, His308, Tyr309) or a salt bridge (Asp156 and Arg229). The conservation of many of these residues in the binding sites of other nucleotide sugar aminotransferases also suggests that they may be crucial for enzyme function.[24,28] In order to explore whether these residues were indeed critical for catalysis, a panel of nine WbpE alanine mutants were generated using site-directed mutagenesis.

In addition to eight residues identified above, a mutant in which the catalytic lysine (Lys185) was replaced with alanine was prepared as well. All mutants were overexpressed and purified to high yield, and then tested for activity using the WbpB/WbpE coupled enzyme assay as previously described.[15] As anticipated, the activity of the Lys185Ala mutant was completely abolished, but the other eight mutants exhibited nearly complete turnover of substrate to product under these reaction conditions, suggesting that the loss of binding affinity of any one of these residues may be compensated for by the cumulative efforts of the others (Figure 4-11). Individual reaction rates were not determined due to the difficulty in isolating the extremely labile ketone substrate of WbpE as well as the need for using the coupled WbpB/WbpE assay,[15] which would mask the effect of a point mutation on the aminotransferase alone.

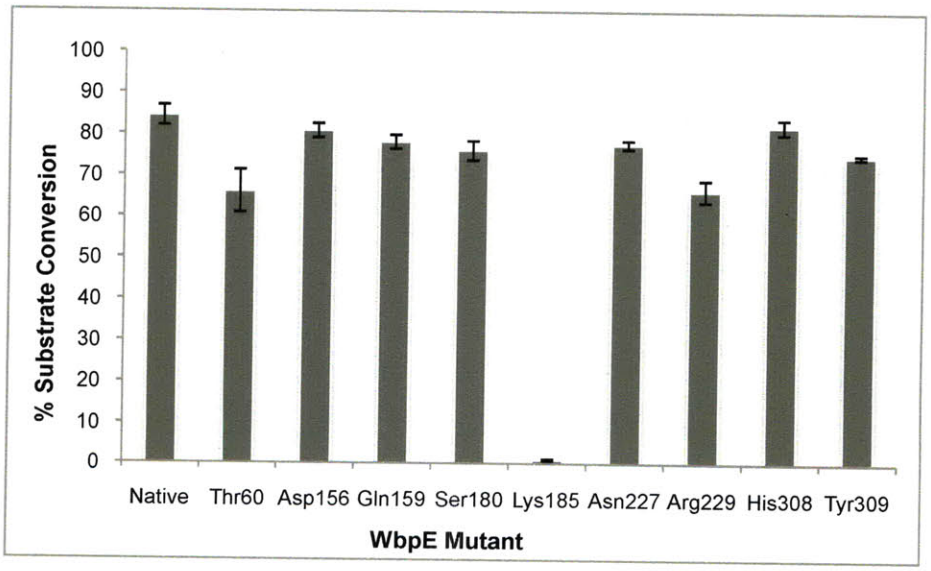


Figure 4-11: Overall activity of the WbpE alanine mutants, as determined through the use of the coupled WbpB/WbpE assay. Crude reaction mixtures were analyzed by capillary electrophoresis to determine total percent of UDP-GlcNAc(3NH_2)A formation; data represent the average of two experiments.

Comparison of WbpE External Aldimine with Other Aminotransferases

To date, there have been three other reported crystal structures of Fold Type I aminotransferases bound to their nucleotide sugar substrates as the external aldimine (Figure 4-12). The first published complex was that of PseC, an aminotransferase from *H. pylori* involved in the biosynthesis of pseudaminic acid.[24] Shortly thereafter, the structures of both DesI, a key enzyme in the *S. venezuelae* desosamine biosynthesis pathway, and QdtB, an aminotransferase required for synthesis of TDP-Qui*p*3NAc in *Thermoanaerobacterium thermosaccharolyticum*, were presented.[25,28] Taken together, these structures provide an opportunity to gain a deeper understanding of the substrate specificity that governs the function of these enzymes.

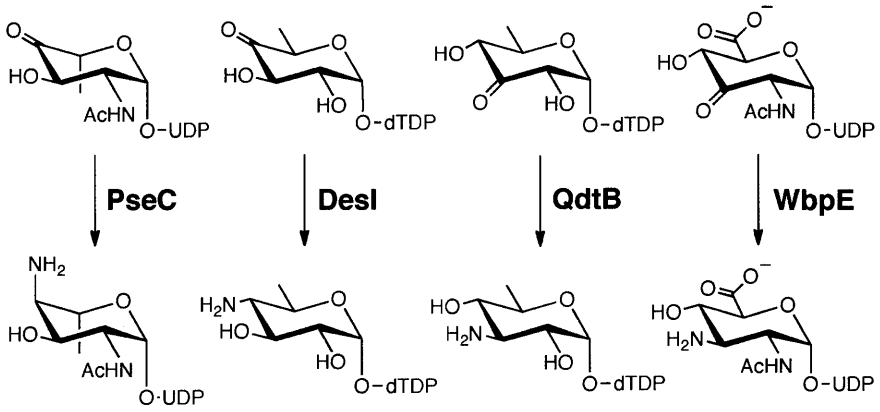


Figure 4-12: Reactions carried out by homologous nucleotide sugar aminotransferases PseC (*H. pylori*), DesI (*S. venezuelae*), QdtB (*T. thermosaccharolyticum*), and WbpE (*P. aeruginosa*). All reactions require PLP and involve the concomitant conversion of L-glutamate to α-ketoglutarate.

A comparison of the active site orientations of the external aldimines in the four known structures (WbpE, PseC, DesI and QdtB) is presented in Figure 4-13. It can be seen that while the ligands are positioned quite differently within the binding sites depending primarily on the functional groups on the hexose rings, the structures do share some common features. As

expected, the PLP cofactors are positioned in the same basic orientation. Also, each of the four hexose substrates contains a hydroxyl group at either the C3" (QdtB and WbpE) or C4" (PseC and DesI) position, and these hydroxyls are all situated on the same side of the active site near the catalytic lysine and the cofactor phosphate or an asparagine, in the case of DesI. Interestingly, all four hexoses seem to be bordered on the top and bottom faces by aromatic side chains, effectively sealing them off from the surrounding environment. One explanation for this observation is that association with these aromatic rings may serve to both stabilize the positively charged reaction intermediates through cation-π interactions and also prevent unwanted water molecules from entering into the active site and interfering with catalysis.

The many differences in nucleotide sugar orientation within the active sites depicted in this analysis provide a glimpse into the requirements for substrate binding and specificity. In general, the hexose rings with smaller functional groups and less polarity are in contact with fewer atoms within the active site. This is rather unexpected, as often substrate binding causes conformational changes within the enzyme active site to create an induced fit. For example, while the nucleotide sugars bound to QdtB and WbpE are both anchored to the cofactor through the C3" position, the QdtB ligand makes far fewer contacts to the active site atoms because of the smaller hexose modifications than the highly functionalized WbpE ligand. Similarly, the C6" hexose carbons of the QdtB, PseC and WbpE ligands are all pointing in the same general direction, yet while there are no atoms within 4 Å of the QdtB and PseC C6" methyl groups, the C6" carboxylate of WbpE is coordinated to both the Arg229 in the β-hairpin domain as well as to a well-ordered water molecule. It is tempting to propose a relationship between increased space around the ligand hexose ring and decreased enzyme specificity. In their study of QdtB, Thoden et al made a preliminary effort to address this theory by showing

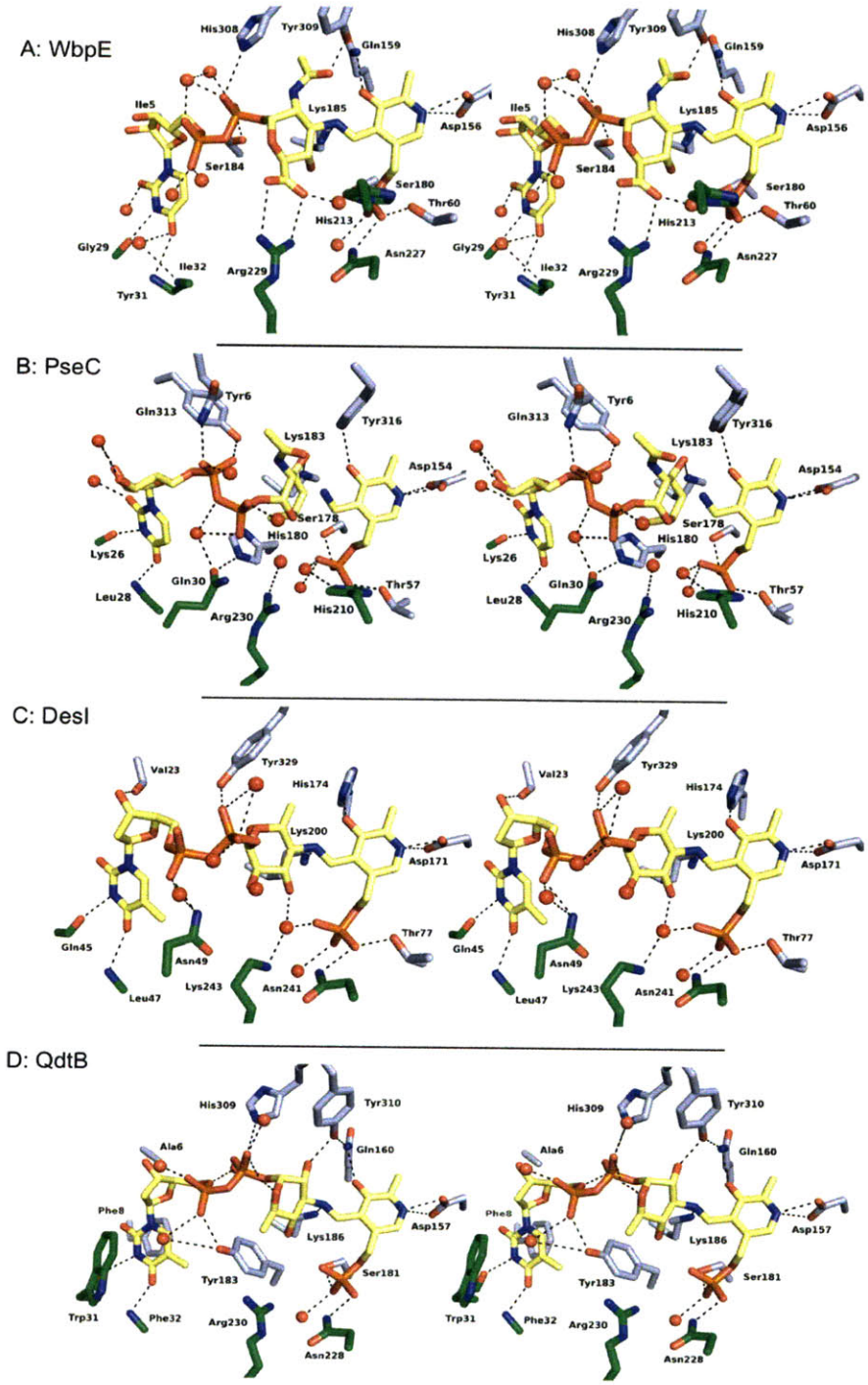


Figure 4-13: Comparison of WbpE (A), PseC (B), DesI (C), and QdtB (D) in complex with the corresponding external aldimines. For each structure, residues from subunit 1 are colored gray and those from subunit 2 are colored green. Interactions within 3.6 Å of the ligand are shown as black lines, and residues lining the top face of the enzyme were removed for the sake of clarity. All four complexes are positioned with the PLP cofactor in the same orientation to highlight the differences in ligand conformation in each active site. In addition, the basic residue important for α-KG binding (Arg or Lys) is labeled (PDB codes: 3NUB, 2FNU, 2PO3, 3FRK).

that QdtB accepts the C4" epimer of its natural substrate, presumably due to the lack of protein contacts between the C4" position of the hexose and the active site.[28] However, further biochemical and biophysical studies are required to explore this hypothesis.

Identification of a Non-Prolyl Cis Amide Bond in the WbpE Ligand Binding Site

A close analysis of the WbpE structure revealed that the enzyme a contains non-prolyl *cis* amide bond between residues His308 and Tyr309 (Figure 4-14). The side chains of both of these residues line the active site of the protein and participate in important contacts with the nucleotide sugar to orient it properly. Studies of the prevalence of non-prolyl *cis* amide bonds within protein structures deposited in the Protein Data Bank indicate that they are extremely rare due to the increased steric strain between the neighboring Cα atoms. Interestingly, the location of these *cis* bonds is often found within functionally important regions of the protein. In addition, more than one-third of the characterized proteins containing *cis* amide bonds are present in carbohydrate-binding proteins.[53] As depicted in Figures 4-6 and 4-13, both His308 and Tyr309 participate in key hydrogen bonds with the external aldimine intermediate, framing the upper region of the active site. However, it is clear that these residues are not absolutely crucial for catalysis, as our studies using alanine mutations at these positions still yielded functional enzyme (Figure 4-11). The close proximity of both His308 and Tyr309 to the hexose and pyrophosphate moieties held in place by the *cis* amide bond suggests that perhaps these residues impart specificity to the enzyme, effectively allowing only substrates of a certain shape to fit in the active site. Further biochemical analysis is required to explore this hypothesis, but difficulty in obtaining the ketone substrate of WbpE complicates these efforts.

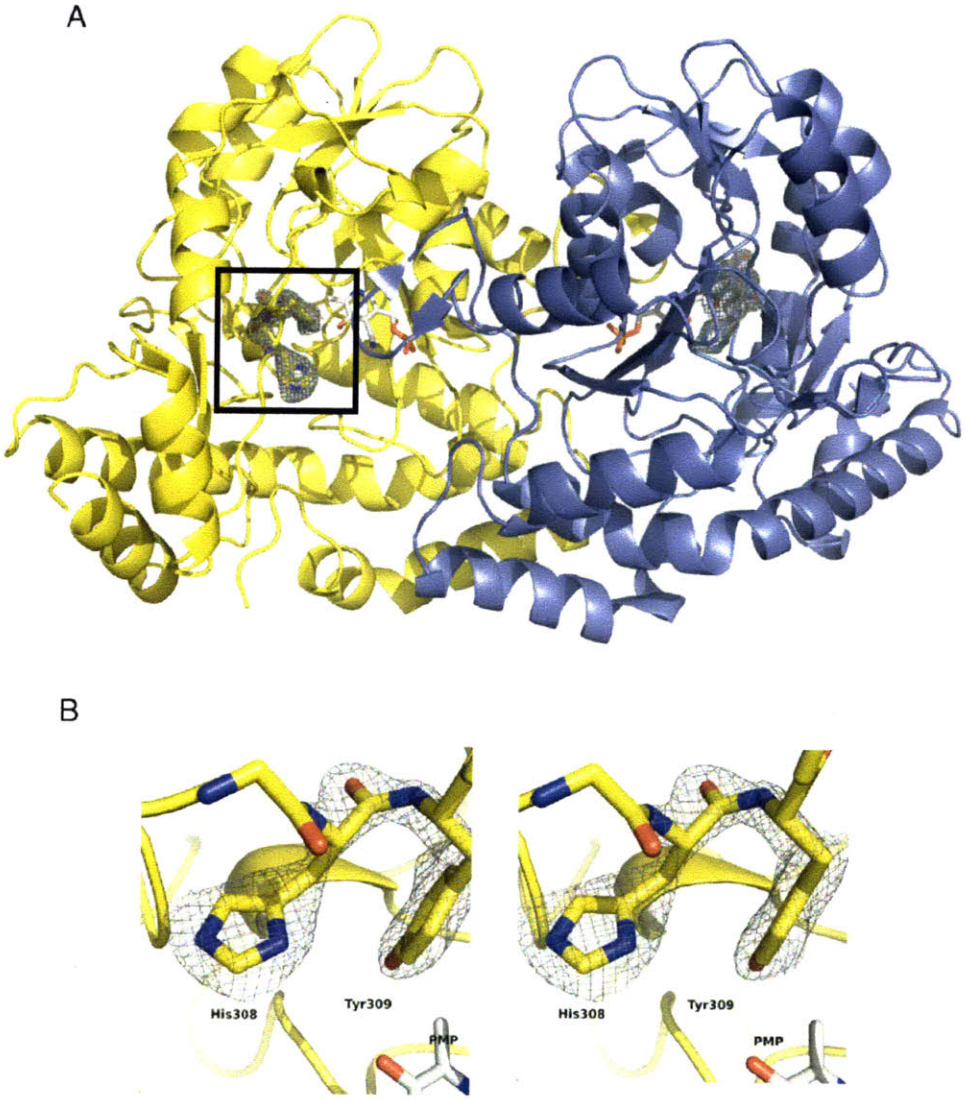


Figure 4-14: The non-prolyl cis amide bond between His308 and Tyr309 in the external aldimine-bound structure of WbpE. (A) The location of the amide bond within the homodimer, where the boxed area indicates the region shown in the stereo figure. (B) A close-up view of the cis amide bond in stereo. The electron density represent a $2(F_o\text{-}F_c)$ map contoured at 2.5σ in which the His and Tyr residues were omitted from the calculation.

Discovery of WbpE Homologs in Related Bacterial Pathogens

While eukaryotes utilize only a limited set of carbohydrates as the basic building blocks for protein glycosylation, prokaryotes routinely incorporate a far greater number of sugar structures in order to maintain the sheer complexity and diversity of their glycans.[54]

Biochemical evidence has suggested that despite this structural diversity, diacetylated aminuronic acids such as ManNAc(3NAc)A are exceedingly rare. This class of highly functionalized sugars has primarily been identified in the complex cell wall matrices of a few pathogenic bacteria, such as *P. aeruginosa* and *B. pertussis*, as well as the flagellar glycoproteins of certain methanogenic archaea.[55,56] However, an analysis of the rapidly growing database of newly sequenced prokaryotic genomes suggests that perhaps these aminuronic acids are more prevalent than previously suspected. A recent search for WbpE homologs resulted in over fifty matches with high sequence identity (> 40%); the eleven closest homologs are presented in Figure 4-15, with a full sequence alignment is provided in Figure 4-16.

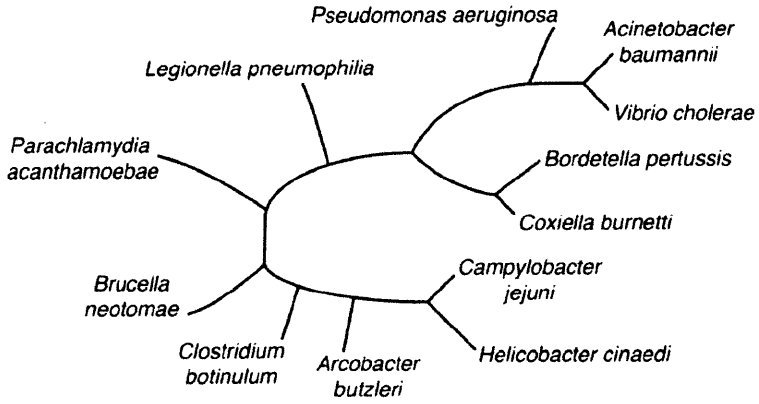


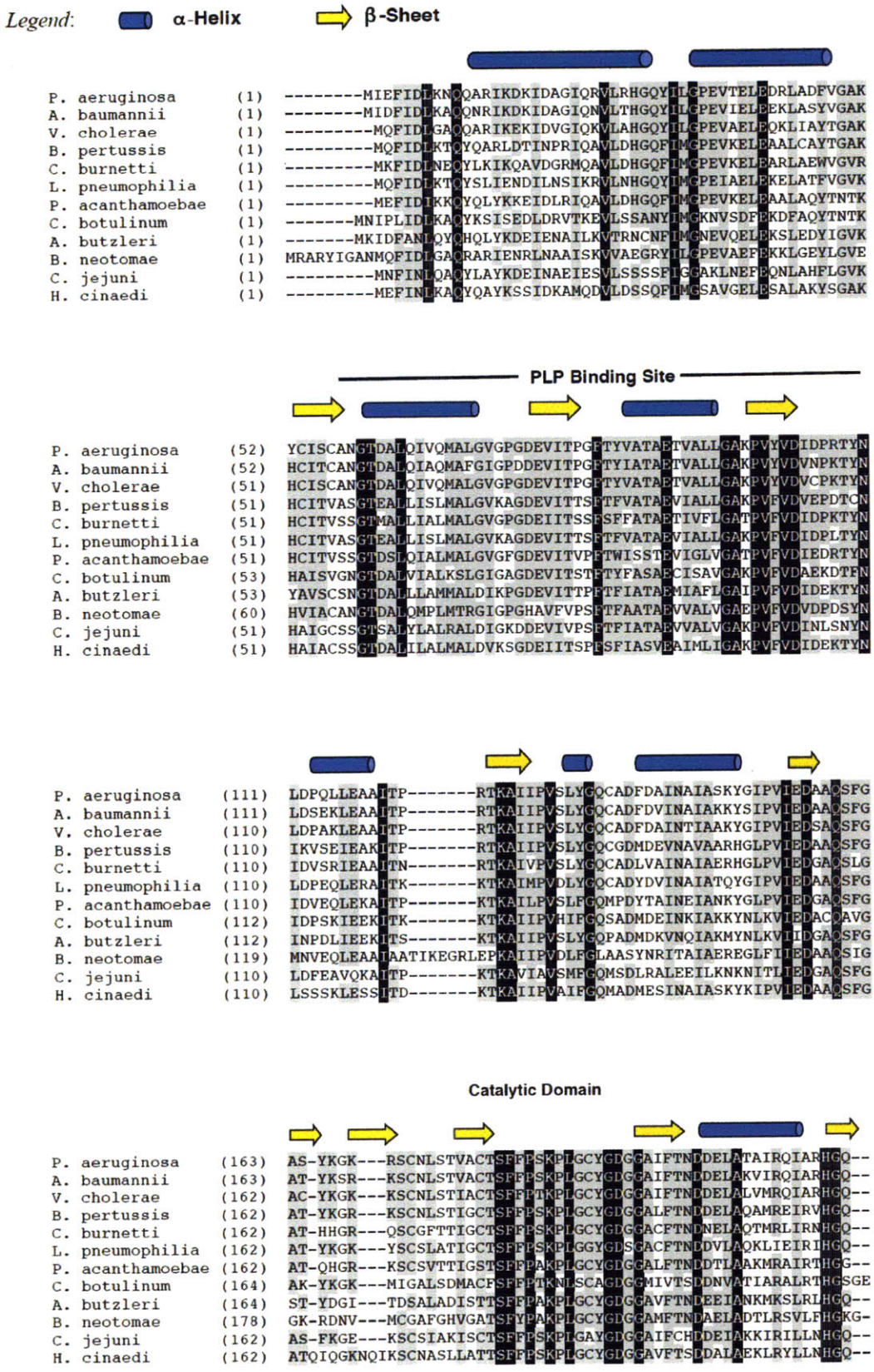
Figure 4-15: Phylogenetic tree indicating homologs of WbpE from pathogenic bacteria. Each homolog has a sequence identity of > 40% with respect to WbpE.

Interestingly, these homologs are all found in pathogenic bacteria, most of which have not been subject to in depth biochemical characterization. The presence of WbpE homologs in these organisms may imply that a ManNAc(3NAc)A-type carbohydrate exists in the cell wall matrix, where it could play a contributing role in modulating virulence and evasion of host defenses, though more work is required to characterize the cell wall components of these organisms to confirm their chemical makeup. Nevertheless, the protein structures reported in

this study may offer insight into the biosynthetic enzymes in other pathogenic organisms and provide a unique opportunity for identification of important new targets for antibacterial drug development.

Efforts Towards Obtaining a Structure of the WbpB/WbpE Complex

As described in Chapter 3, recent work has revealed that the dehydrogenase WbpB and aminotransferase WbpE work together to catalyze formation of UDP-GlcNAc($3NH_2$)A from UDP-GlcNAcA using a novel NAD^+ recycling mechanism. In this pathway, the α-KG product of WbpE is shunted back to WbpB, where it is reduced to 2-hydroxyglutarate (2-HG) as a means to regenerate NAD^+ from NADH, thus preparing WbpB for another round of catalysis. The underlying reason for why the activities of these two enzymes are coupled together is currently unclear. One hypothesis is that the UDP-GlcNAc(3keto)A intermediate is labile, and thus coordinating the activities of WbpB and WbpE serves to bring these two enzymes into close physical proximity to limit the exposure of ketone to the surrounding aqueous environment. However, as described in Chapter 3, the presence of both enzymes is not required for WbpB catalysis, as addition of α-KG to the WbpB reaction mixture was sufficient to drive product formation (Figure 3-10). Current work entails collaboration with the Allen laboratory (Boston University) to obtain a crystal structure of a complex containing both WbpB and WbpE. Analysis of this structure could uncover the nature of the interaction between these two enzymes that enables efficient transfer of the UDP-GlcNAc(3keto)A intermediate and α-KG.


Legend:
α-Helix
β-Sheet
PLP Binding Site
Catalytic Domain
P. aeruginosa (1) --------MIEFIDLKNQQARIKDKIDAGIQRVLRHGQYILGPEVTELEDRLADFVGAK
A. baumannii (1) --------MIDFIDLKAQQNRIKDKIDAGIQNVLTHGQYILGPEVIELEKLASYVGAK
V. cholerae (1) ---------MQFIDLGAQQARIKEKIDVGIQKVLAHGQYILGPEVAELEQKLIAYTGAK
B. pertussis (1) ---------MQFIDLKTQYQARLDTINPRIQAVLDHGQFIMGPEVKELEAALCAYTGAK
C. burnetti (1) ---------MKFIDLNEQYLKIKQAVDGRMQAVLDHGQFIMGPEVKELEARLAEWVGVR
L. pneumophilia (1) ---------MQFIDLKTQYSLIENDILNSIKRVLNHGQYIMGPEIAELEKELATFVGVK
P. acanthamoebae (1) ---------MEFIDLKKQYQLYKKEIDLRIQAVLDHGQFIMGPEVKELEAALAQYTNTK
C. botulinum (1) -------MNIPLIDLKAQYKSISEDLDRVTKEVLSSANYIMGKNVSDFEKDFAQYTNTK
A. butzleri (1) -------MKIDFANLQYQHQLYKDEIENAILKVTRNCNFIMGNEVQELEKSLEDYIGVK
B. neotomae (1) MRARYIGANMQFIDLGAQRARIENRLNAAISKVVAEGRYILGPEVAEFEKKLGEYLGVE
C. jejuni (1) ---------MNFINLQAQYLAYKDEINAEIESVLSSSSFILGGAKLNEPEQNLAHFLGVK
H. cinaedi (1) ---------MEFINLKAQYQAYKSSIDKAMQDVLDSSQFIMGSAVGELESALAKYSGAK
P. aeruginosa (52) YCISCANGTDALQIVQMALGVGPGDEVITPGFTYVATAETVALLGAKPVYVDIDPRTYN
A. baumannii (52) HCITCANGTDALQIAQMAFGIGPDDEVITPGFTYIATAETVALLGAKPVYVDVNPKTYN
V. cholerae (51) HCISCANGTDALQIVQMALGVGPGDEVITPGFTYVATAETVALLGAKPVYVDVCPKTYN
B. pertussis (51) HCITVASGTEALLISLMALGVKAGDEVITTSFTFVATAEVIALLGAKPVFVDVEPDTCN
C. burnetti (51) HCITVSSGTMALLIALMALGVGPGDEIITSSFSFFATAETIVFLGATPVFVDIDPKTYN
L. pneumophilia (51) HCITVASGTEALLISLMALGVKAGDEVITTSFTFVATAEVIALLGAKPVFVDIDPLTYN
P. acanthamoebae (51) HCITVSSGTDSLQIALMALGVGFGDEVITVPFTWISSTFVIGLVGATPVFVDIEDRTYN
C. botulinum (53) HAISVGNGTDALVIALKSLGIGAGDEVITSTFTYFASAECISAVGAKPVFVDAEKDTFN
A. butzleri (53) YAVSCSNGTDALLLAMMALDIKPGDEVITTPFTFIATAEMIAFLGAIPVFVDIDEKTYN
B. neotomae (60) HVIACANGTDALQMPLMTRGIGPGHAVFVPSFTFAATAEVVALVGAEPVFVDVDPDSYN
C. jejuni (51) HAIGCSSGTSALYLALRALDIGKDDEVIVPSFTFIATAEVVALVGAKPVFVDINLSNYN
H. cinaedi (51) HAIACSSGTDALILALMALDVKSGDEIITSPFSFIASVEAIMLIGAKPVFVDIDEKTYN
P. aeruginosa (111) LDPQLLEAAITP-------RTKAIIPVSLYGQCADFDAINAIASKYGIPVIEDAAQSFG
A. baumannii (111) LDSEKLEAAITP-------RTKAIIPVSLYGQCADFDVINAIAKKYSIPVIEDAAQSFG
V. cholerae (110) LDPAKLEAAITP-------RTKAIIPVSLYGQCADFDAINTIAAKYGIPVIEDSAQSFG
B. pertussis (110) IKVSEIEAKITP-------RTKAIIPVSLYGQCGDMDEVNAVAARHGLPVIEDAAQSFG
C. burnetti (110) IDVSRIEAAITN-------RTKAIVPVSLYGQCADLVAINAIAERHGLPVIEDGAQSLG
L. pneumophilia (110) LDPEQLERAITK-------KTKAIMPVDLYGQCADYDVINAIATQYGIPVIEDAAQSFG
P. acanthamoebae (110) IDVEQLEKAITP-------KTKAILPVSLFGQMPDYTAINEIANKYGLPVIEDGAQSFG
C. botulinum (112) IDPSKIEEKITK-------KTKAIIPVHIFGQSADMDEINKIAKKYNLKVIEDACQAVG
A. butzleri (112) INPDLIEEKITS-------KTKAIIPVSLYGQPADMDKVNQIAKMYNLKVIEDGAQSFG
B. neotomae (119) MNVEQLEAAIAATIKEGRLEPKAIIPVDLFGLAASYNRITAIAEREGLFIIEDAAQSIG
C. jejuni (110) LDFEAVQKAITP-------KTKAVIAVSMFGQMSDLRALEEILKNKNITLIEDGAQSFG
H. cinaedi (110) LSSSKLESSITD-------KTKAIIPVAIFGQMADMESINAIASKYKIPVIEDAAQSFG
P. aeruginosa (163) AS-YKGK---RSCNLSTVACTSFFPSKPLGCYGDGGAIFTNDDELATAIRQIARHGQ--
A. baumannii (163) AT-YKSR---KSCNLSTVACTSFFPSKPLGCYGDGGAIFTNDDELAKVIRQIARHGQ--
V. cholerae (162) AC-YKGK---KSCNLSTIACTSFFPTKPLGCYGDGGAIFTNDDELALVMRQIARHGQ--
B. pertussis (162) AT-YKGR---KSCNLSTIGCTSFFPSKPLGCYGDGGALFTNDDELAQAMREIRVHGQ--
C. burnetti (162) AT-HHGR---QSCGFTTIGCTSFFPSKPLGCYGDGGACFTNDNELAQTMRLIRNHGQ--
L. pneumophilia (162) AT-YKGK---YSCSLATIGCTSFFPSKPLGGYGDSGACFTNDDVLAQKLIEIRIHGQ--
P. acanthamoebae (162) AT-QHGR---KSCSVTTIGSTSFFPAKPLGCYGDGGALFTNDDTLAAKMRAIRTHGG--
C. botulinum (164) AK-YKGK---MIGALSDMACFSFFPTKNLSCAGDGGMIVTSDDNVATIARALRTHGSGE
A. butzleri (164) ST-YDGI---TDSALADISTTSFFPAKPLGCYGDGGAVFTNDEEIANKMKSLRIHGQ--
B. neotomae (178) GK-RDNV---MCGAFGHVGATSFYPAKPLGCYGDGGAMFTNDAELADTLRSVLFHGKG-
C. jejuni (162) AS-FKGE---KSCSIAKISCTSFFPSKPLGAYGDGGAIFCHDDEIAKKIRILLNHGQ--
H. cinaedi (162) ATQIQGKNQIKSCNASLLATTSFFPSKPLGCYGDGGAVFTSDDALAEKLRYLLNHGQ--

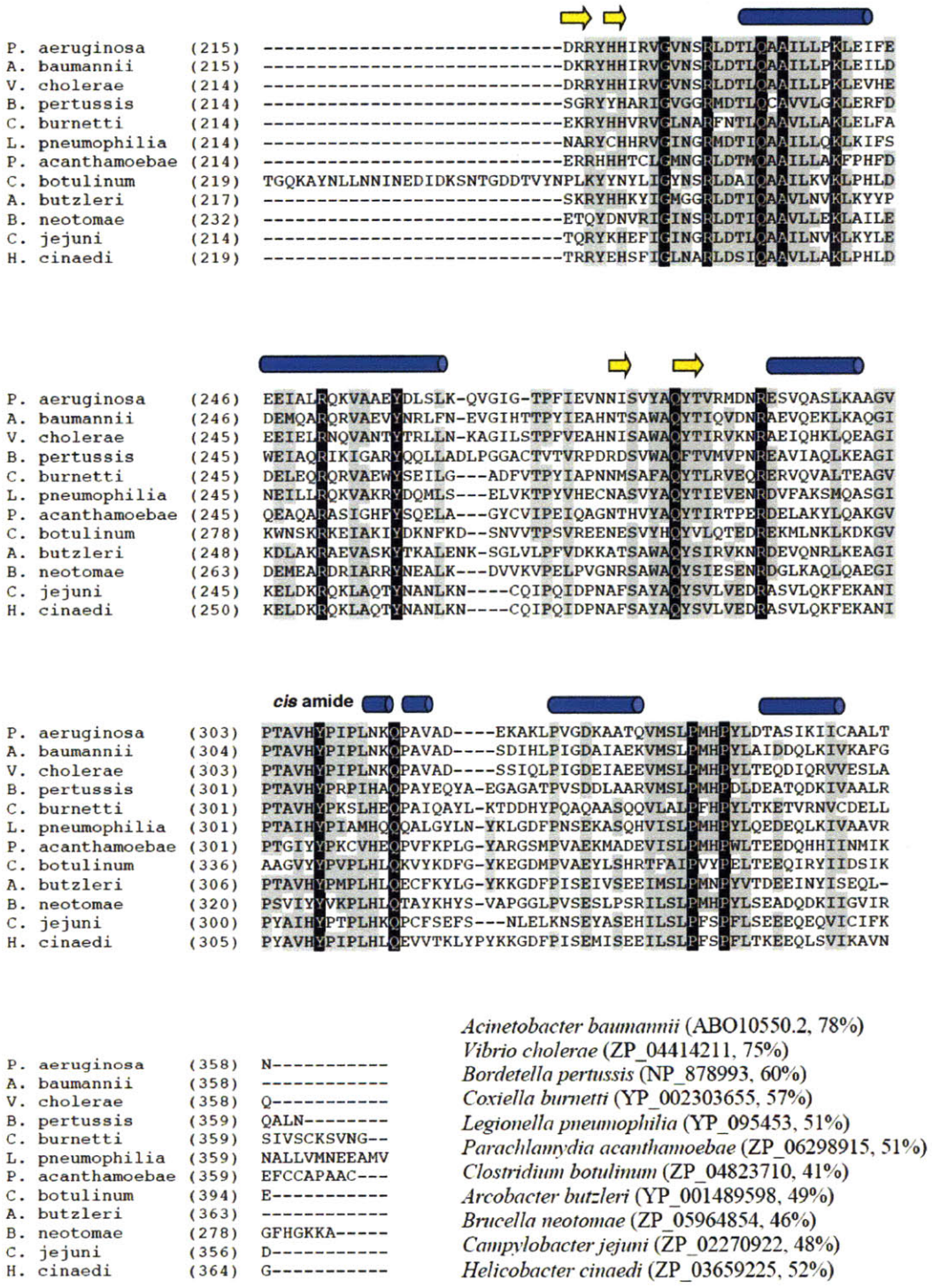


Figure 4-16: Sequence alignment of WbpE with closely related homologs from pathogenic bacteria. Homologs were obtained using BLAST, and the alignment performed with ClustalW. Conserved residues are highlighted in black and similar residues are indicated in gray. Secondary structure was determined from analysis of the PMP-bound WbpE structure using PROCHECK. The GenBank accession numbers of each homolog are listed, along with sequence identity to WbpE.

Conclusions

In this chapter, the structure of WbpE in complex with three different ligands, PMP, PLP, and the UDP-GlcNAc($3NH_2$)A product as the external aldimine, is presented. These studies describe the first structure of an aminotransferase that binds a nucleotide sugar modified at the C2", C3", and C6" positions, and details the means by which such a highly functionalized substrate is accommodated in the enzyme active site. In addition, the dimeric structure of WbpE observed in the crystal structure is confirmed to also exist in solution by sedimentation velocity analytical ultracentrifugation. In order to achieve a broad perspective of substrate specificity among this class of aminotransferases, the external aldimine structure is compared to that of homologous transferases bound to their substrates; this analysis reveals that highly functionalized nucleotide sugar substrates make more contacts with their respective enzyme active sites, which may have an implication on substrate specificity. Future work in this area will focus on gaining structural insight into the interactions between WbpB and WbpE that facilitate the novel NAD^+ recycling behavior outlined in Chapter 3.

Acknowledgements

I am grateful to the staff at beamline X6A at the National Synchrotron Light Source, particularly Dr. Jean Jankoncic for assistance with data collection and analysis. Special thanks to Debby Pheasant of the Biophysical Instrumentation Facility at MIT for assistance with analytical centrifugation, Dr. Robert Grant of the Biology Department and members of the Drennan laboratory, particularly Dr. Christine Phillips and Yan Kung, for technical advice, and Dr. Matthieu Sainlos for encouragement during synchrotron trips. I am also deeply indebted to Dr. Nelson Olivier for his guidance and support.

Experimental Methods

General Methods

Unless otherwise noted, all reagents were obtained commercially and used without further purification. Oligonucleotides were purchased from Sigma Life Sciences (St. Louis, MO), and restriction endonucleases were obtained from New England Biolabs (Ipswich, MA). Sequencing of all bacterial plasmids was performed by the MIT CCR Biopolymers Laboratory (Cambridge, MA). Amino acids, including L-SeMet, were purchased from Sigma Alrich (St. Louis, MO). Crystal screening kits were obtained from Hampton Research (Aliso Viejo, CA) and Emerald BioSystems (Bainbridge Island, WA). Crystal trays and other crystallography equipment were purchased from Hampton Research. MALDI-TOF mass spectrometry was performed on the Voyager system (Applied Biosystems), and capillary electrophoresis was carried out on the P/ACE MDQ instrument (Beckman Coulter) as detailed in Chapter 3.

Cloning of WbpE

The *wbpE* gene was amplified from *Pseudomonas aeruginosa* PAO1-LAC genomic DNA (ATCC) by the polymerase chain reaction as previously described, using primers to introduce 5' *Bam*HI and 3' *Xho*I restriction sites that are outlined in the Appendix (Table 1).[15] The resulting oligonucleotide was then inserted into a modified pET32a vector (Novagen) using standard molecular biology techniques. Site-directed mutagenesis was performed using the QuikChange protocol from Stratagene with the *wbpE*-pET32a plasmid as a template. All constructs yielded proteins with an N-terminal His_6-tag followed by a TEV site for tag removal.

Overexpression of WbpE

The *wbpE*-pET32a plasmid was transformed into *E. coli* BL21-CodonPlus(DE3) RIL competent cells (Stratagene) for heterologous expression, using both kanamycin (50 μg/mL) and chloramphenicol (30 μg/mL) for selection. The cell culture was grown to an optical density (600 nm) of 0.8-1.0 at 37 °C in Luria-Bertani broth; the culture was subsequently cooled to 16 °C and protein expression was induced through the addition of IPTG (1mM). After 16 h, the cells were harvested by centrifugation (5000*g*) and the resulting cell pellets were stored at -80 °C for future use.

Incorporation of Selenomethionine Into WbpE

Incorporation of selenomethionine was accomplished using the method of metabolic inhibition described elsewhere with slight modification.[38] Briefly, 0.5 L of M9 medium supplemented with 1 mM $MgSO_4$, 3 mM $FeSO_4$, 0.4% (w/v) glucose, 0.5% (w/v) thiamine, kanamycin (50 μg/mL) and chloramphenicol (30 μg/mL) was inoculated with 5 mL starter culture and allowed to incubate at 37 °C until the desired optical density (0.8-1.0) was obtained. Prior to induction, the following amino acids were added to the flask and the culture was incubated for 20 minutes to allow for the inhibition of methionine biosynthesis: L-lysine (100 mg/L), L-phenylalanine (100 mg/L), L-threonine (100 mg/L), L-isoleucine (50 mg/L), L-leucine (50 mg/L), L-valine (50 mg/L), L-selenomethionine (60 mg/L). After addition of IPTG (1 mM) to induce protein expression, the culture was handled as described above.

Purification of WbpE

All steps were performed at 4 °C. WbpE was purified from cell pellets using Ni-NTA resin (Qiagen) as previously described.[15] After overnight dialysis to remove the imidazole and lower the salt concentration, the N-terminal His_6-tag was removed by proteolysis with TEV over the course of three days while stirring in dialysis buffer (50 mM HEPES, pH 8.0/100 mM NaCl/4% glycerol/0.5 mM EDTA/5 mM DTT); the removal of the tag was confirmed by Western blot analysis using an Anti-His_4 anitbody (Qiagen). The protein was then subjected to size-exclusion chromatography using a Superdex 200 16/60 column (GE Healthcare) in running buffer composed of 25 mM HEPES, pH 8.0/100 mM NaCl/0.5% glycerol. Fractions containing monodispersed protein were pooled and analyzed by SDS-PAGE for purity and MALDI mass spectrometry to quantify selenomethionine incorporation. Purified protein was routinely utilized within 24 hrs to prevent aggregation.

Sedimentation Velocity Analytical Ultracentrifugation

Experiments were conducted in an Optima XL-I ultracentrifuge (Beckman Coulter) using an An60 Ti rotor at 4 °C with a rotor speed of 42,000 rpm. Data were acquired by monitoring absorbance at 280 nm through quartz cell windows. A sample of WbpE (36 μM) was dialyzed against 25 mM HEPES, pH 8.0/100 mM NaCl/0.5% glycerol for 24 h prior to the experiment; the centrifugation run also included dialysis buffer as a blank. Data were analyzed with the software package SEDNANAL[57] to determine the oligomeric state of the protein in solution.

Crystallization of WbpE

Prior to setting up crystal trays, WbpE was concentrated to 10 mg/mL in the size-exclusion running buffer. For crystallization in the presence of ligands, PLP and/or UDP-GlcNAc($3NH_2$)A were added to the protein solution at a final concentration of 500 μM and 10 mM respectively and allowed to incubate on ice for 1 h. Crystals were obtained at 25 °C from a hanging drop by mixing 1.5 μL protein solution with 1.5 μL reservoir solution. All crystals except for the SeMet derivative were grown in a reservoir solution containing 0.1 M Bis-Tris, pH 5.5, 0.2 M ammonium sulfate and 25% PEG 3350. SeMet crystals were grown in a reservoir solution containing 0.1 M Bis-Tris, pH 5.5, 0.2 M ammonium acetate, 10 mM $SrCl_2$ and 25% PEG 3350. Crystals were cryoprotected in the corresponding reservoir solution supplemented with 20% glycerol and substrate as necessary.

Data Collection

Diffraction data were collected on beamline X6A (National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY) at 110 K, using an Si(111) channel cut monochromator and a toroidal focusing mirror. Reflections were recorded with an ADSC Quantum 270 detector, with an active area of 270 mm^2 and a pixel size of 65 μm in a 4084 x 4084 array. For the SeMet derivative, data were collected at the selenium peak (12667 eV), absorption edge (12660 eV) and remote energies (12867 eV). Data sets were indexed and scaled using HKL2000, and the scaled intensities were converted to structure factors using the program TRUNCATE in the CCP4 suite of programs.[40,41] Data collection parameters are summarized in Table 1.

Structure Determination and Refinement

The structure of the WbpE-SeMet derivative was solved using the method of multiwavelength anomalous diffraction.[34] Using data collected at the selenium peak and truncated to 2.5 Å, the program SOLVE was employed to locate the heavy atom sites and generate experimental phases.[32] A total of four out of five selenium atoms were located for each protein subunit. The initial model was built using the prime-and-switch phasing feature of RESOLVE to minimize model bias.[42,43] The PMP-bound structure was solved by molecular replacement using Phaser with the SeMet derivative as the initial search model.[44] Further model building and refinement were carried out using Coot and Refmac.[45,46] Five percent of the data were used to calculate the R_{free} values for cross-validation of the refinement process.[47] The structures of the PLP and external aldimine complexes were subsequently solved by molecular replacement in the same manner, using the completed PMP-bound structure with the cofactor removed as the input model. Water molecules were added using both ARP/wARP and Coot, and ligands were modeled into each structure after the R_{free} value was below 30%.[48] All refined structures were validated using PROCHECK, SFCHECK, and MolProbity,[49-51] and the final refinement statistics are presented in Table 1. All molecular images were generated using PyMOL.[58]

Functional Characterization of WbpE Alanine Mutants

Enzyme mutants were analyzed for function using the coupled WbpB/WbpE assay slightly modified from that described elsewhere.[15] In brief, 2.5 μg of each WbpE mutant was incubated with UDP-GlcNAcA (0.75 mM), NAD^+ (0.2 mM), L-glutamate (25 mM), PLP (0.1 mM), DTT (2.5 mM), $MgCl_2$ (2 mM) and 2.5 μg of WbpB in 60 μL of 50 mM HEPES buffer (pH 8.0) for 2 h at 30 °C. Capillary electrophoresis (P/ACE MDQ System, Beckman Coulter) was used to

monitor enzyme activity as outlined previously, and the presence of product was confirmed by the addition of starting material into the reaction mixtures and the observation of a new peak on the CE chromatogram.

References

1. Pier, G.B. *Pseudomonas aeruginosa* lipopolysaccharide: A major virulence factor, initiator of inflammation and target for effective immunity. *Int. J. Med. Microbiol.* **2007**, *297*, 277-295.
2. Obritsch, M.D.; Fish, D.N.; MacLaren, R.; Rose, J. Nosocomial infections due to multidrug-resistant *Pseudomonas aeruginosa*: epidemiology and treatment options. *Pharmacotherapy* **2005**, *25*, 1353-1364.
3. Tang, H.B.; DiMango, E.; Bryan, R.; Gambello, M.; Iglewski, B.H.; Goldberg, J.B.; Prince, A. Contribution of specific *Pseudomonas aeruginosa* virulence factors to pathogenesis of pneumonia in a neonatal mouse model of infection. *Infect. Immun.* **1996**, *64*, 37-43.
4. Diaz, M.H.; Shaver, C.M.; King, J.D.; Musunuri, S.; Kazzaz, J.A.; Hauser, A.R. *Pseudomonas aeruginosa* induces localized immunosuppression during pneumonia. *Infect. Immun.* **2008**, *76*, 4414-4421.
5. Son, M.S.; Matthews, W.J., Jr.; Kang, Y.; Nguyen, D.T.; Hoang, T.T. In vivo evidence of *Pseudomonas aeruginosa* nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. *Infect. Immun.* **2007**, *75*, 5313-5324.
6. Levy, S.B.; Marshall, B. Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.* **2004**, *10*, S122-S129.
7. Mesaros, N.; Nordmann, P.; Plesiat, P.; Roussel-Delvallez, M.; Van Eldere, J.; Y., G.; Van Laethem, Y.; Jacobs, F.; Lebecque, P.; Malfroot, A.; Tulkens, P.M.; Van Bambeke, F. *Pseudomonas aeruginosa*: resistance and therapeutic options at the turn of the new millennium. *Clin. Microbiol. Infect.* **2007**, *13*, 560-578.
8. Hancock, R.E. Resistance mechanisms in *Pseudomonas aeruginosa* and other nonfermative gram-negative bacteria. *Clin. Infect. Dis.* **1998**, *27*, S93-S99.
9. Cryz, S.J., Jr.; Pitt, T.L.; Furer, E.; Germanier, R. Role of lipopolysaccharide in virulence of *Pseudomonas aeruginosa*. *Infect. Immun.* **1984**, *44*, 508-513.
10. Hancock, R.E.; Mutharia, L.M.; Chan, L.; Darveau, R.P.; Speert, D.P.; Pier, G.B. *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis: a class of serum-sensitive, nontypable strains deficient in lipopolysaccharide O side chains. *Infect. Immun.* **1983**, *42*, 170-177.
11. Engels, W.; Endert, J.; Kamps, M.A.; van Boven, C.P. Role of lipopolysaccharide in opsonization and phagocytosis of *Pseudomonas aeruginosa*. *Infect. Immun.* **1985**, *49*, 182-189.
12. Dasgupta, T.; de Kievit, T.R.; Masoud, H.; Altman, E.; Richards, J.C.; Sadovskaya, I.; Speert, D.P.; Lam, J.S. Characterization of lipopolysaccharide-deficient mutants of *Pseudomonas aeruginosa* derived from serotypes O3, O5, and O6. *Infect. Immun.* **1994**, *62*, 809-817.
13. Kochetkov, N.K.; Knirel, Y.A. Structure of lipopolysaccharides from gram-negative bacteria. III. Structure of O-specific polysaccharides. *Biochemistry (Moscow)* **1994**, *59*, 1325-1383.
14. Rocchetta, H.L.; Burrows, L.L.; Lam, J.S. Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 523-553.

15. Larkin, A.; Imperiali, B. Biosynthesis of UDP-GlcNAc(3NAc)A by WbpB, WbpE, and WbpD: enzymes in the Wbp pathway responsible for O-antigen assembly in *Pseudomonas aeruginosa* PAO1. *Biochemistry* **2009**, *48*, 5446-5455.
16. Burrows, L.L.; Charter, D.F.; Lam, J.S. Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PAO1) B-band lipopolysaccharide gene cluster. *Mol. Microbiol.* **1996**, *22*, 481-495.
17. John, R.A. Pyridoxal phosphare-dependent enzymes. *Biochim. Biophys. Acta* **1995**, *1248*, 81-96.
18. Eliot, A.C.; Kirsch, J.F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **2004**, *73*, 383-415.
19. Ford, G.C.; Eichele, G.; Jansonius, J.N. Three-dimensional structure of a pyridoxal-phosphate-dependent enzyme, mitochondrial aspartate aminotransferase. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1559-2563.
20. Tiniakos, D.G.; Vos, M.B.; Brunt, E.M. Nonalcoholic fatty liver disease: pathology and pathogenesis. *Annu Rev Pathol* **2010**, *5*, 145-171.
21. Neuschwander, B.A.; Ünalp, A.; Creer, M.H. Influence of local reference populations on upper limits of normal for serum alanine aminotransferase levels. *Arch Intern Med* **2008**, *168*, 663-666.
22. Jansonius, J.N. Structure, evolution and action of vitamin B_6-dependent enzymes. *Curr. Opin. Struct. Biol.* **1998**, *8*, 759-769.
23. Noland, B.W.; Newman, J.M.; Hendle, J.; Badger, J.; Christopher, J.A.; Tresser, J.; Rutter, M.E.; Sanderson, W.E.; Müller-Dieckmann, H.-J.; Gajiwala, K.S.; Buchanan, S.G. Structural studies of *Salmonella typhimurium* ArnB (PmrH) aminotransferase: a 4-amino-4-deoxy-L-arabinose lipopolysaccharide-modifying enzyme. *Structure* **2002**, *10*, 1569-1580.
24. Schoenhofen, I.C.; Lunin, V.V.; Julien, J.-P.; Li, Y.; Ajamian, E.; Matte, A.; Cygler, M.; Brisson, J.-R.; Aubry, A.; Logan, S.M.; Bhatia, S.; Wakarchuk, W.W.; Young, N.M. Structural and functional characterization of PseC, an aminotransferase involved in the biosynthesis of pseudaminic acid, an essential flagellar modification in *Helicobacter pylori*. *J. Biol. Chem.* **2006**, *281*, 8907-8916.
25. Burgie, E.S.; Holden, H.M. Molecular architecture of DesI: a key enzyme in the biosynthesis of desosamine. *Biochemistry* **2007**, *46*, 8999-9006.
26. Burgie, E.S.; Thoden, J.B.; Holden, H.M. Molecular architecture of DesV from *Streptomyces venezuelae*: A PLP-dependent transaminase involved in the biosynthesis of the unusual sugar desosamine. *Protein Sci* **2007**, *16*, 887-896.
27. Cook, P.D.; Holden, H.M. GDP-perosamine synthase: structural analysis and production of a novel trideoxysugar. *Biochemistry* **2008**, *47*, 2833-2840.
28. Thoden, J.B.; Schäffer, C.; Messner, P.; Holden, H.M. Structural analysis of QdtB, an aminotransferase required for the biosynthesis of dTDP-3-acetamido-3,6-dideoxy-α-D-glucose. *Biochemistry* **2009**, *48*, 1553-1561.
29. Pasternak, A.; White, A.; Jeffery, C.J.; Medina, N.; Cahoon, M.; Ringe, D.; Hedstrom, L. The energetic cost of induced fit catalysis: Crystal structures of trypsinogen mutants with enhanced activity and inhibitor affinity. *Protein Sci.* **2001**, *10*, 1331-1342.
30. Rould, M.A. Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. *Methods in Enzymology* **1997**, *276*, 461-472.

31. Ke, H. Overview of isomorphous replacement phasing. *Methods in Enzymology* **1997**, *276*.
32. Terwilliger, T.C.; Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *D55*, 849-861.
33. Sheldrick, G.M. A short history of SHELX. *Acta Crystollagr., Sect. A: Found. Crystallogr.* **2008**, *A64*, 112-122.
34. Hendrickson, W.A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **1991**, *254*, 51-58.
35. Hendrickson, W.A.; Horton, J.R.; LeMaster, D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **1990**, *9*, 1665-1672.
36. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods in Enzymology* **1997**, *276*, 523-530.
37. Guerrero, S.A.; Hecht, H.J.; Hofmann, B.; Biebl, H.; Singh, M. Production of selemethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl. Microbiol. Biotechnol.* **2001**, *56*, 718-723.
38. Van Duyne, G.D.; Standaert, R.F.; Karplus, P.A.; Schreiber, S.L.; Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **1993**, *29*, 105-124.
39. Benson, T.E.; Filman, D.J.; Walsh, C.T.; Hogle, J.M. An enzyme-substrate complex involved in bacterial cell wall biosynthesis. *Nat. Struct. Biol.* **1995**, *2*, 644-653.
40. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **1997**, *276*, 307-326.
41. Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *D50*, 760-763.
42. Terwilliger, T.C. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 38-44.
43. Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56*, 965-972.
44. Storoni, L.C.; McCoy, A.J.; Read, R.J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *D60*, 432-438.
45. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *D60*, 2126-2132.
46. Murshudov, G.N.; Vagin, A.A.; Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *D53*, 240-255.
47. Brünger, A.T. Assessment of phase accuracy by cross validation: the free *R* value, methods and applications. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1993**, *D49*, 24-36.
48. Lamzin, V.S.; Perrakis, A.; Wilson, K.S., The ARP/wARP suite for automated construction and refinement of protein models. In Int. Tables for Crystallography, Rossmann, M. G.; Arnold, E., Eds. Kluwer Academic Publishers: Dordrecht, 2001; Vol. Vol F: Crystallography of biological macromolecules, pp 720-722.

49. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **1993**, *26*, 283-291.
50. Vaguine, A.A.; Richelle, J.; Wodak, S.J. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *D55*, 191-205.
51. Davis, I.W.; Leaver-Fay, A.; Chen, V.B.; Block, J.N.; Kapral, G.J.; Wang, G.J.; Wang, X.; Murray, L.W.; Arendal III, B.; Snoeyink, J.; Richardson, J.S.; Richardson, D.C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375-W383.
52. Okamoto, A.; Ishii, S.; Hirotsu, K.; Kagamiyama, H. The active site of *Paracoccus denitrificans* aromatic amino acid aminotransferase has contrary properties: flexibility and rigidity. *Biochemistry* **1999**, *38*, 1176-1184.
53. Jabs, A.; Weiss, M.; Hilgenfeld. Non-proline Cis peptide bonds in proteins. *J. Mol. Biol.* **1999**, *286*, 291-304.
54. Varki, A.; Cummings, R.D.; Freeze, H.H.; Stanley, P.; Bertozzi, C.R.; Hart, G.W.; Etzler, M.E. *Essentials of Glycobiology*; 2nd Ed.; Cold Spring Harbor Laboratory Press: New York, 2008.
55. Kelly, J.; Logan, S.M.; Jarrell, K.F.; VanDyke, D.J.; Vinogradov, E.V. A novel *N*-linked flagellar glycan from *Methanococcus maripaludis*. *Carb. Res.* **2009**, *344*, 648-653.
56. Caroff, M.; Brisson, J.-R.; Martin, A.; Karibian, D. Structure of the *Bordetella pertussis* 1414 endotoxin. *FEBS Lett.* **2000**, *477*, 8-14.
57. Stafford, W.F.; Sherwood, P.J. Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. *Biophys. Chem.* **2004**, *108*, 231-243.
58. Delano, W.L., The PyMOL molecular graphics system. DeLano Scientific: San Carlos, CA, 2002.

Chapter 5 Efforts Towards the Development of Dolichyl-Linked Disaccharide Substrates for the Study of Archaeal OTases

Introduction

The vast complexity of oligosaccharide structures found in nature is derived in part from the enzymatic action of glycosyltransferases, which catalyze glycan assembly through the addition of individual carbohydrate units.[1] Although the most common glycosyl substrates of these enzymes are nucleotide diphosphate sugars (UDP-Glc, UDP-Gal, GDP-Man), glycosyltransferases recognize a wide variety of other activated carbohydrate substrates as well, including nucleotide monophosphate sugars (CMP-NeuAc), lipid phosphates (Dol-P-Man), and unsubstituted phosphates.[2] A recent survey of the CAZy database indicates that over 12,500 unique glycosyltransferases have been identified; these proteins are further classified into 91 subfamilies based on sequence analysis, making glycosyltransferases one of the most numerous enzyme classes in nature.[3,4]

Although the function of glycosyltransferases was originally described in the early 1970s, difficulties with protein expression and purification have complicated efforts to examine these enzymes in great detail.[1,5] The first X-ray crystal structure of a glycosyltransferase was reported in 1994 for the bacteriophage T4-glucosyltransferase;[6] to date, over 100 X-ray crystal structures of glycosyltransferases representing 23 distinct subfamilies have been deposited in the Protein Data Bank (PDB). Interestingly, a comparative analysis of these structures indicates that while these enzymes do not exhibit high sequence homology, they feature a surprising level of structural homology such that only two main structural folds have been identified, termed GT-A and GT-B.[2] Both glycosyltransferase types are composed of $\alpha/\beta/\alpha$ sandwiches, with GT-A members containing a single Rossmann fold where the active site is located between two β-sheets, and GT-B types comprising two Rossmann folds that surround the catalytic active site (Figure 5-1).[5] In addition, glycosyltransferases that do not fit into either category have been

recently described, such as the CstII sialyltransferase from *Campylobacter jejuni*;[7] however, these classifications have not been verified biochemically. The conserved structural homology among known glycosyltransferases suggests that a majority of these enzymes evolved from a small number of progenitor sequences.[2] In contrast, glycosidases, which are responsible for the hydrolysis of glycosyl bonds, have been shown to adopt a wide variety of folds, including all α, all β, or mixed α/β structures, and thus cannot be classified into a small number of structural classes.[5,8]

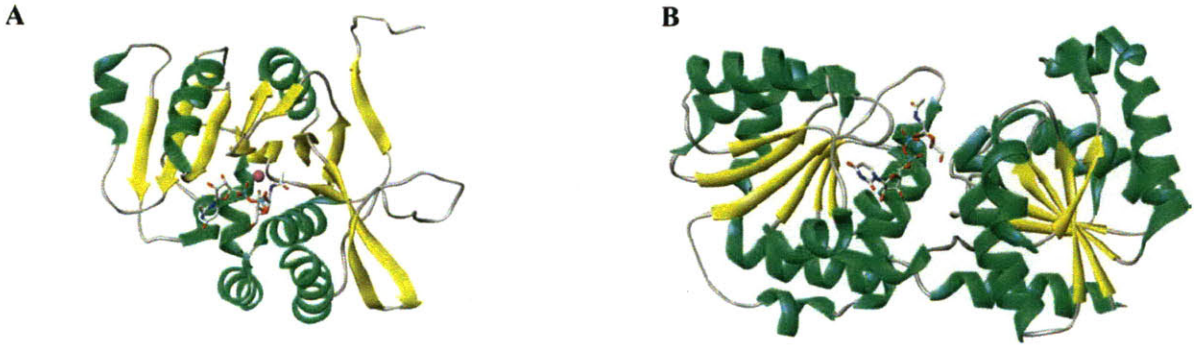


Figure 5-1: Representative structures of the two major glycosyltransferase folds. (A) GT-A fold, mouse EXTL2 α-1,4-*N*-acetylhexosaminyl transferase complexed with UDP-GalNAc (PDB code 1OMZ); (B) GT-B fold, *Escherichia coli* MurG β-1,4 *N*-acetylglucosamine transferase complexed with UDP-GlcNAc (PDB code 1NLM). Figure modified from Breton et al.[5]

In recent years, a great deal of progress has been made towards understanding the catalytic mechanism of glycosyltransferases. In general, these enzymes are categorized as either inverting or retaining, based upon the stereochemistry of the newly formed bond relative to the anomeric center of the glycosyl donor (Figure 5-2).[9] Based on biochemical studies, the mechanism of inverting glycosyltransferases is proposed to involve an S_N2-type displacement of the activating group of the glycosyl donor by a nucleophilic moiety within the glycosyl

acceptor, facilitated by an active site base. In contrast, the mechanism of retaining glycosyltransferases is less clear, as both double displacement and S_N1-type reaction mechanisms have been suggested.[2] Further mechanistic studies are required in order to determine how these enzymes achieve retention of the glycosyl donor configuration.

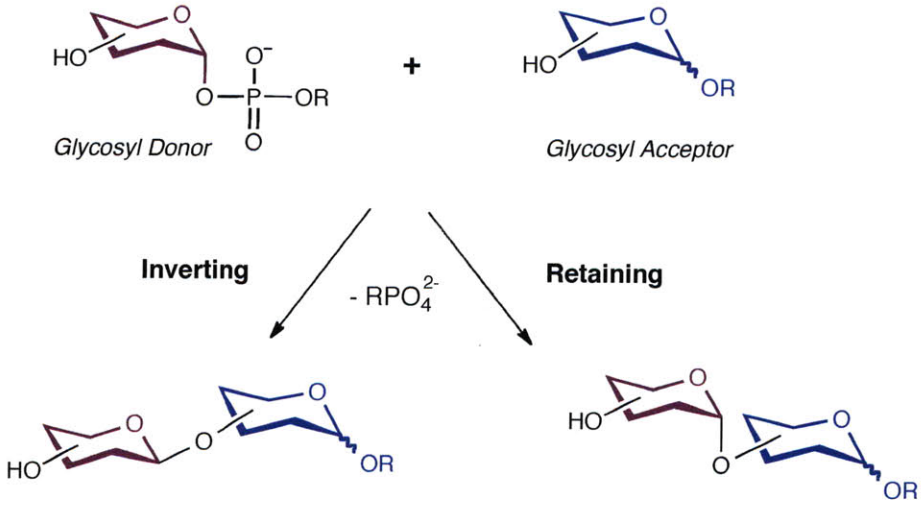


Figure 5-2: General scheme depicting the stereochemical outcomes of the two main functional classes of glycosyltransferases, inverting and retaining.[2]

As outlined in Chapter 1, the biosynthesis of asparagine-linked glycans in all three kingdoms of life requires the action of many glycosyltransferase for oligosaccharide assembly. One of the most interesting aspects of this process is the formation of the polyisoprenyl-linked disaccharide, which is the product of the second glycosyltransferase in the pathway (Figure 5-3). In the yeast *Saccharomyces cerevisiae*, two proteins (Alg13 and Alg14) have been shown to be required for formation of the Dol-PP-$GlcNAc_2$ disaccharide. A combination of biochemical and structural studies has revealed that the C-terminal domains of both Alg13 and Alg14 interact to

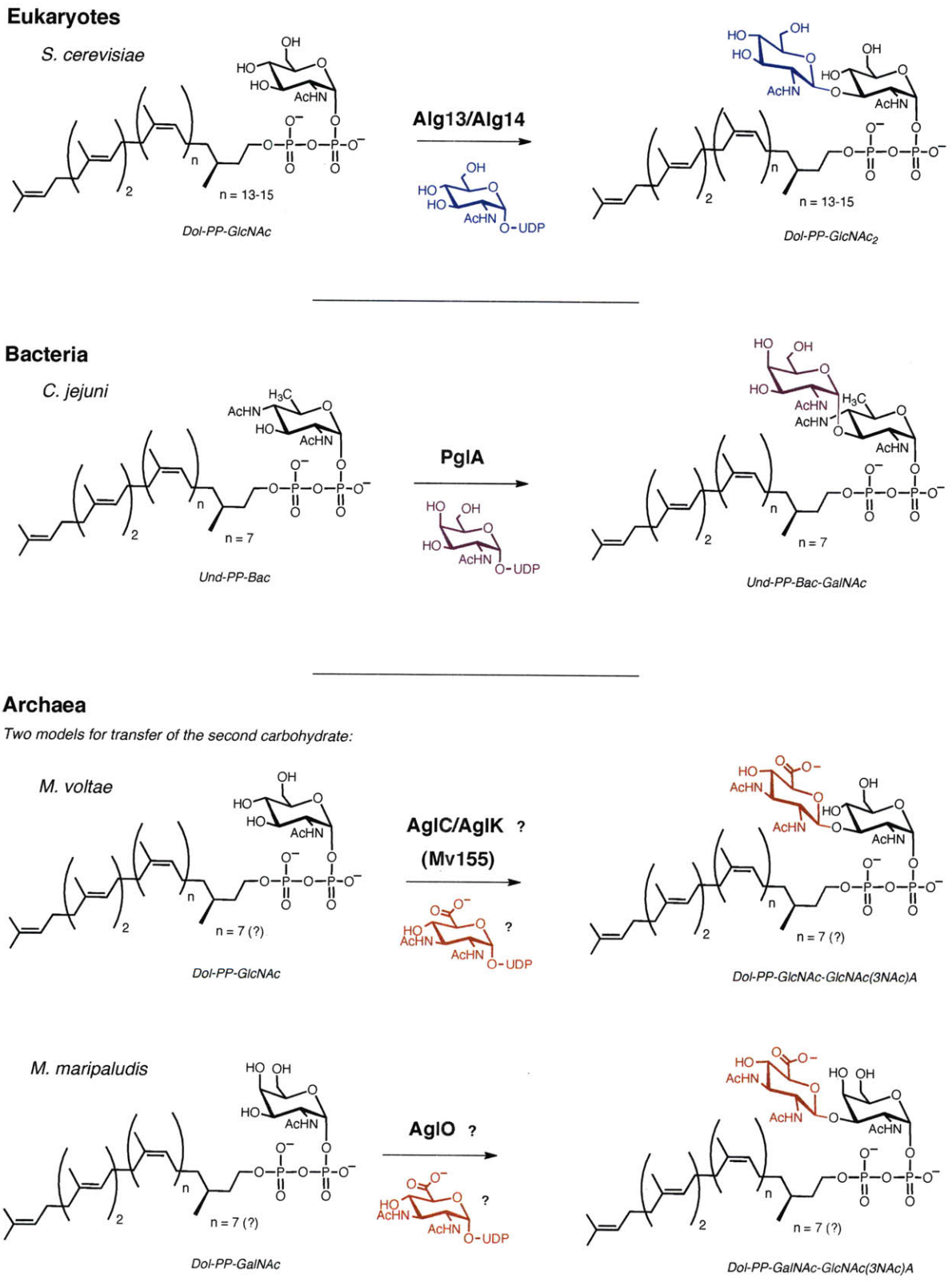


Figure 5-3: Comparison of polyisoprenyl-linked disaccharide formation in *N*-linked glycan biosynthesis across all three kingdoms of life. The two models proposed in the archaea *M. voltae* and *M. maripaludis* have not yet been subject to biochemical verification.

form a functional glycosyltransferase capable of catalyzing GlcNAc transfer.[10,11] In contrast, the analogous reaction in the bacterium *Campylobacter jejuni* has been shown to require only the single protein, PglA.[12] Although the steps involved in the biosynthesis of the dolichyl-linked disaccharide in archaea have not yet been biochemically characterized, early genetic studies have proposed two different models for this reaction. In the first model described in *Methanococcus voltae*, two proteins (AglC and AglK) are implicated in this role, similar to the pathway in eukaryotes.[13] In the second model, genetic studies suggest that disaccharide formation is carried out by a single glycosyltransferase in *Methanococcus maripaludis*, AglO, which is reminiscent of the bacterial glycosylation system.[14] As these two methanogens are very closely related, it seems unlikely both models are in fact true; further studies are required to determine which (if any) model accounts for disaccharide formation in the archaeal kingdom.

The overall focus of the work in this thesis is directed at gaining a deeper understanding of the last step of *N*-linked glycan biosynthesis, namely the reaction catalyzed by the oligosaccharyl transferase (OTase). Early work toward this goal involved the identification of a highly-expressing OTase from the archaeon *M. voltae*, which was produced in *E. coli* at a 100-fold improvement over the bacterial OTase PglB from *C. jejuni* (Chapter 1). In order to begin to characterize the function of this OTase, a suitable dolichyl-linked disaccharide substrate is required (Figure 5-4). To this end, we have achieved the chemoenzymatic synthesis of the unique nucleotide sugar UDP-GlcNAc(3NAc)A using a series of enzymes in the Wbp pathway of the opportunistic pathogen *Pseudomonas aeruginosa*, outlined in Chapter 3. In this chapter, efforts towards coupling this sugar to a dolichyl-linked monosaccharide in order to generate the desired disaccharide substrates are described. Four different glycosyltransferases from the archaea *M. voltae* and *M. maripaludis* that have been proposed to catalyze this reaction were

cloned, overexpressed, and evaluated for function. Despite testing these enzymes in a wide variety of reaction conditions, we have been unable to observe transfer of the GlcNAc(3NAc)A donor onto the dolichyl-linked monosaccharide acceptor. The synthesis of a suite of dolichyl-linked monosaccharides to further probe these putative glycosyltransferases is also described, in which the size of the dolichol and the identity of the monosaccharide is varied. Ultimately, these novel dolichyl-linked disaccharides will aid in detailed biochemical studies of the mechanism of oligosaccharide transfer, which has been a long-standing interest of the Imperiali laboratory.

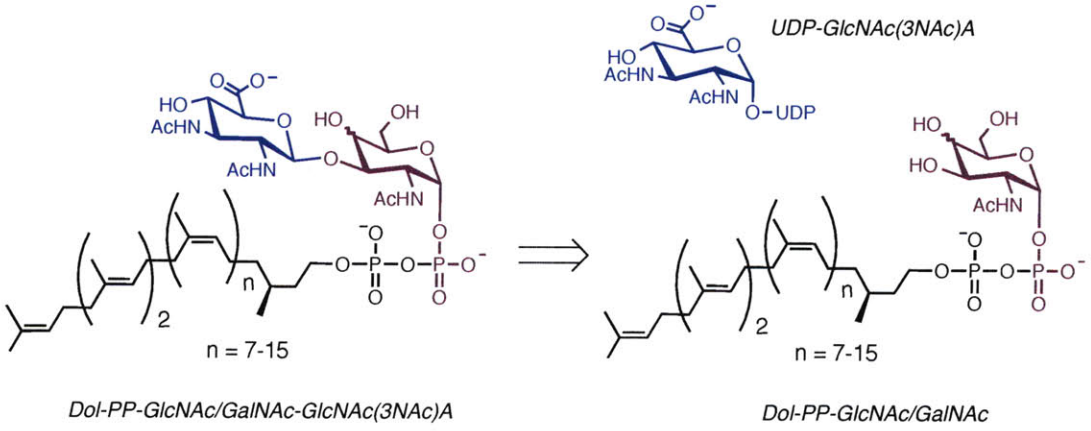


Figure 5-4: Retrosynthetic strategy for the generation of the desired dolichyl-linked disaccharides for the study of the archaeal OTases, Dol-PP-GlcNAc-GlcNAc(3NAc)A or Dol-PP-GalNAc-GlcNAc(3NAc)A.

Results and Discussion

Characterization of Putative GlcNAc(3NAc)A Glycosyltransferases

In order to generate the desired dolichyl-linked disaccharide substrates Dol-PP-GlcNAc-GlcNAc(3NAc)A and Dol-PP-GalNAc-GlcNAc(3NAc)A for study of the *M. voltae* and *M. maripaludis* OTases, a general method for coupling GlcNAc(3NAc)A to Dol-PP-GlcNAc or

Dol-PP-GalNAc was required. While this reaction could be achieved using chemical synthesis, the activation of either the GlcNAc(3NAc)A donor or the dolichylphosphate monosaccharide acceptor entails multiple steps mainly involving protecting group manipulation and was predicted to be low-yielding; therefore, a chemoenzymatic method was pursued. In this approach, the desired β-1,3 linkage between the two monosaccharide residues could be installed by an appropriate enzyme without the use of protecting groups. In addition, it was envisioned that this method could be scaled up to generate preparative quantities of the final product for use in further studies.

Over the past four years, the Jarrell laboratory has identified four possible glycosyltransferases that may be responsible for the transfer of a GlcNAc(3NAc)A donor onto either Dol-PP-GlcNAc or Dol-PP-GalNAc using a series of genetic knockout studies; these proteins are summarized in Table 5-1. Three of these glycosyltransferases, Mv155, AglC, and AglK, are from *M. voltae*, and the fourth, AglO, is found in *M. maripaludis*. Due to problems with genetic manipulation of *M. voltae*, the Jarrell lab has made several different claims as to the identity of the second glycosyltransferase required in *N*-linked glycan biosynthesis. First, the Mv155 protein was implicated as the GlcNAc(3NAc)A transferase;[15] however, two years later, this sequence was retracted from the NCBI database and replaced with AglC.[16] In 2009, the group reported that the transfer of GlcNAc(3NAc)A required AglK in addition to AglC, although biochemical evidence for this proposal is still lacking. Also in 2009, the genes responsible for *N*-linked glycan biosynthesis in the more genetically tractable yet closely related methanogen *M. maripaludis* were reported. In this organism, it appeared that only a single gene, *aglO*, was required to add the GlcNAc(3NAc)A to the growing dolichyl-linked glycan in contrast to the proposed involvement of both *aglC* and *aglK* in *M. voltae*. In general, all four

putative GlcNAc(3NAc)A transferases identified are predicted to be in the GT-2 subfamily and to exhibit the GT-A structural fold based on protein sequence. In comparison, the second glycosyltransferase required in both the yeast *S. cerevisiae* and the bacterium *C. jejuni* are members of the GT-1 subfamily and thus are characterized by GT-B type structure. A summary of these glycosyltransferases is shown in Table 5-1, and the predicted topologies of these proteins using the TMHMM server (ca.expasy.org/tools) are depicted in Figure 5-5.[17]

Glycosyltransferase	**MW (kDa)**	**GT family**	**GT fold type**
M. voltae AglC	37.3	GT family 2	GT-A
M. voltae AglK	28.2	GT family 2	GT-A
M. voltae Mv155	41.0	GT family 2	GT-A
M. maripaludis AglO	32.4	GT family 2	GT -A
S. cerevisiae Alg13	126.1	GT family 1	GT-B
S. cerevisiae Alg14	24.2	GT family 1	GT-B
C. jejuni PglA	42.7	GT family 1	GT-B

Table 5-1: Summary of proteins implicated in GlcNAc(3NAc)A transfer (AglC, AglK, Mv155, and AglO) compared with Alg13/14 from *S. cerevisiae* and PglA from *C. jejuni*, which have been shown to be responsible for polyisoprenyl-linked disaccharide formation in eukaryotes and bacteria, respectively.

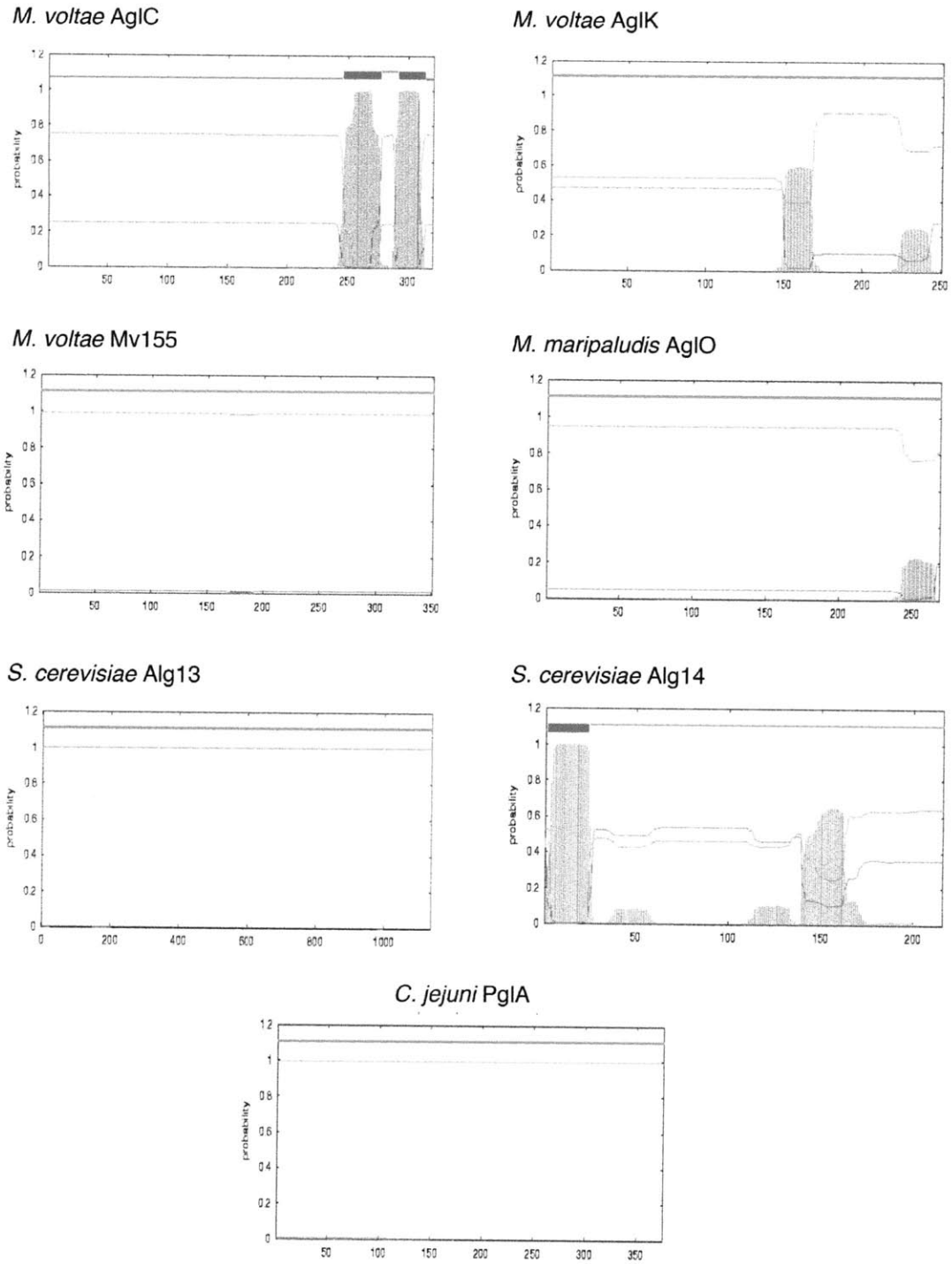


Figure 5-5: Predicted topology of the putative GlcNAc(3NAc)A transferases AglC, AglK, Mv155, and AglO from *M. voltae* and *M. maripaludis*, as well as Alg13 and Alg14 from *S. cerevisiae* and PglA from *C. jejuni*. Topology predictions were generated using the TMHMM server.[17]

In order to test these proteins for GlcNAc(3NAc)A transferase activity, each gene was chemically synthesized by a commercial vendor to contain optimal codons for bacterial expression, then inserted into one of three vectors for protein expression: pET24a(+), which yields a final protein containing an N-terminal T7 tag and a C-terminal His_6 tag; pGEX, which results in a N-terminal glutathione S-transferase (GST) fusion protein; and pGBH, providing a final protein that contains an N-terminal GB1 tag and a C-terminal His_6 tag. The resulting plasmids were transformed into two different *E. coli* competent cell lines to screen for optimal protein expression. As each putative glycosyltransferase is predicted to either contain transmembrane domains or to associate with the cell membrane through hydrophobic regions on the protein surface, all four proteins were prepared as crude membrane fractions or purified using the His_6 or GST tags and compared by Western blot analysis. It appeared that the pET24a(+) constructs expressed in *E. coli* BL21(DE3) RIL cells gave the highest overall yields for each glycosyltransferase (Figure 5-6). In addition, the AglC protein appeared to associate with the cell membranes, while AglK was found in both the membrane and soluble fractions. Mv155 and AglO behaved as soluble proteins.

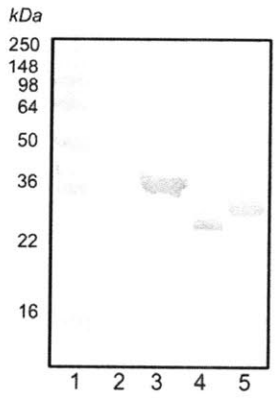


Figure 5-6: Anti-T7 Western blot of the putative GlcNAc(3NAc)A transferases described in this chapter. (1) MW standard; (2) Mv155; (3) AglC; (4) AglK; (5) AglO. AglC was prepared as a crude membrane fraction, while the other proteins were purified using Ni-NTA as outlined.

After purification (or isolation of the crude membrane fraction in the case of AglC), each protein was then tested for GlcNAc(3NAc)A transferase activity. The standard assay for function involved the incubation of Dol-PP-GlcNAc and UDP-[^{3}H]GlcNAc(3NAc)A with the enzyme in question for 2 hrs, after which the reaction was quenched and subjected to an aqueous/organic phase separation. Analysis of the organic layer by scintillation counting indicated whether the radiolabeled GlcNAc(3NAc)A was transferred from the aqueous phase into the organic phase. A large panel containing over 60 different reactions conditions, in which variables such as reaction temperature (4, 25, 30, 37, 65 °C) and pH (6.5, 7.0, 7.5, 8.0, 8.5) were altered and additives including detergents (Triton X-100, DDM, CHAPSO, LDAO), salts (NaCl, KCl, $(NH_4)_2SO_4$), and metals ($MnCl_2$, $MgCl_2$, $CaCl_2$) was utilized to screen each protein. Despite this exhaustive screen, transfer of GlcNAc(3NAc)A to Dol-PP-GlcNAc was not observed. Interestingly, in the case of AglO, a low level of activity (~8%) was detected, despite the use of Dol-PP-GlcNAc as the glycan acceptor instead of the native Dol-PP-GalNAc. However, the poor turnover exhibited by AglO was too low to be of preparative use.

It is unclear why we were unable to observe glycosyltransferase activity. One hypothesis is that these proteins may have exquisite specificity for their dolichyl-linked monosaccharide substrates, and that the C85-C95 Dol-PP-GlcNAc employed in the initial screen of conditions is not utilized by these enzymes as the glycosyl acceptor. An early study in the archaeon *Halobacterium halobium* suggested that the C55-C60 form of dolichol was recognized as the polyisoprenol carrier in archaea, as opposed to the longer C85-C95 dolichols found in yeast and higher eukaryotes.[18] In addition, it may be possible that these proteins have been misannotated. A recent BLAST search using various glycosyltransferases in the *S. cerevisiae N*-linked glycosylation pathway as a probe indicated that both AglC and AglK bear more similarities with

dolichyl-phosphoglucose synthase (DPG, Alg5), the enzyme required for formation of the Dol-P-Glc glycosyl donors in the ER lumen, than with Alg13/14.[19] Further work in this area is required to determine whether these proteins catalyze GlcNAc(3NAc)A transfer, and what substrates are required for this reaction.

Synthesis of Dolichyl-Linked Monosaccharides for Further Study

In an attempt to further characterize the glycosyltransferases described in this study, the synthesis of a small panel of four dolichyl-linked monosaccharides was pursued (Figure 5-7). The size of the dolichyl moiety was altered through the use of the shorter C55 form. In addition, the identity of the monosaccharide was varied by employing either GlcNAc or GalNAc.

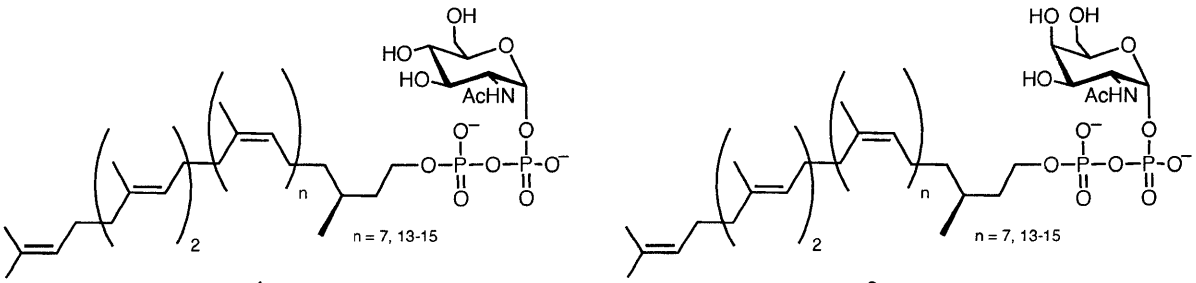


Figure 5-7: Desired dolichyldiphosphate-linked monosaccharides.

The synthesis of these dolichyl-linked monosaccharides involved the coupling of either GlcNAc-1-P or GalNAc-1-P to both forms of dolichylphosphate, C55 or the C85-C95 mixture. As a first step in this synthetic scheme, the phosphorylation of the C85-C95 dolichol was carried

out using phosphoramidite chemistry as previously reported (Figure 5-8);[20,21] the C55 dolichol was obtained as the dolichylphosphate from Professor E. Swiezewska (Polish Academy of Sciences, Warsaw, Poland).

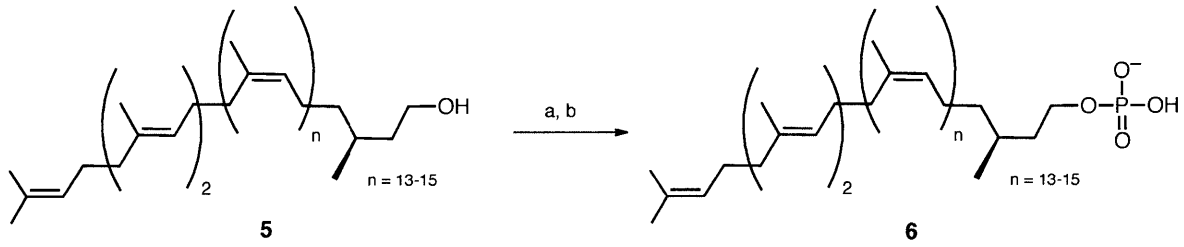


Figure 5-8: Chemical phosphorylation of dolichol. Reagents and conditions: (a) Bis-(2-cyanoethyl)-*N,N'*-diisopropyl phosphoramidite, tetrazole, dry THF, rt; (b) 1% H_2O_2, -78 °C, then 1% NaOMe/MeOH.

The synthesis of the sugar phosphates is outlined in Figure 5-9, and began with the global acetylation of D-glucosamine (**7**) and D-galactoasmine (**8**) with acetic anyhydride in pyridine to afford the tetraacetates. The anomeric hydroxyl group of each sugar was then selectively deprotected by treatment with dimethylamine to give **9** and **10** as a mixture of both the α and β anomers. Incubation of the triacetylated sugars with dibenzyl *N,N'*-phosphoramidite and 1,2,4-triazole resulted in the desired phosphites, which were oxidized to the corresponding protected phosphates **11** and **12** through exposure to a 1% solution of H_2O_2. Deprotection of the sugar phosphates was achieved through hydrogenation in the presence of $Pd(OH)_2/C$. The sugar phosphates **13** and **14** were then activated by incubation with 1,1'-carbodiimidazole (CDI), followed by coupling to both the long (C85-C95) and short (C55) forms of dolichol to yield the desired dolichyl-linked monosaccharides **1**, **2**, **3**, and **4**. Screening of the putative

glycosyltransferases described in this chapter with the panel of dolichyl-linked monosaccharides can now be conducted in an effort to generate the desired Dol-PP-GlcNAc-GlcNAc(3NAc)A or Dol-PP-GalNAc-GlcNAc(3NAc)A, which will then be used to probe the archaeal OTases from *M. voltae* and *M. maripaludis* for function.

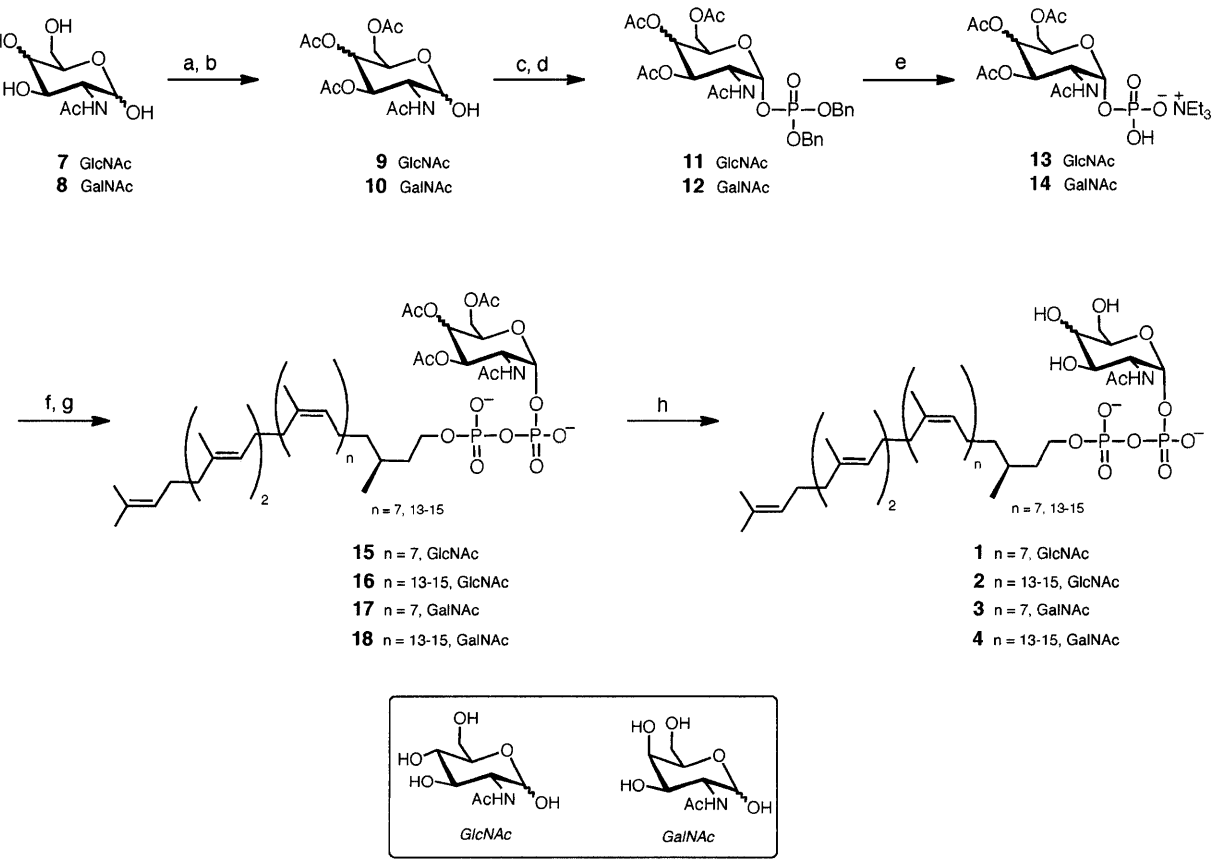


Figure 5-9: Synthesis of dolichyl-linked monosaccharide. Reagents and conditions: (a) Ac_2O, pyr; (b) $HN(CH_3)_2$, CH_3CN; (c) dibenzyl *N,N*-phosphoramidite, 1,2,4-triazole, CH_2Cl_2; (d) 1% H_2O_2, -78 °C; (e) $Pd(OH)_2/C$, H_2, NEt_3; (f) CDI, DMF; (g) MeOH, then Dol-P (C55 or C85-C95, CH_2Cl_2; (g) 1% NaOMe/MeOH.

Conclusions

In this chapter, efforts towards the synthesis of a dolichyl-linked disaccharide substrate for the in depth study of an archaeal OTase are summarized. Earlier work involved the chemoenzymatic synthesis of the unique nucleotide sugar UDP-GlcNAc(3NAc)A using a series of enzymes in the Wbp pathway of the opportunistic pathogen *Pseudomonas aeruginosa* (Chapter 3). In an attempt to couple this sugar to dolichylphosphate (C85-C95), four different glycosyltransferases from the archaea *M. voltae* and *M. maripaludis* that have been implicated in this reaction through genetic experiments were cloned, overexpressed, and evaluated for function. Despite testing these enzymes in a wide variety of reaction conditions, we have been unable to observe transfer of the GlcNAc(3NAc)A donor onto the dolichyl-linked monosaccharide acceptor. To broaden the scope of the polyisoprenyl acceptor substrates used to probe these putative glycosyltransferases, the synthesis of a suite of dolichyl-linked monosaccharides was also described, in which the size of the polyisoprene and the identity of the monosaccharide are varied. Future work will involve completion of these dolichyl-linked disaccharide substrates, which ultimately will aid in detailed biochemical studies of the mechanism of oligosaccharide transfer that has been a long-standing interest of the Imperiali laboratory.

Acknowledgements

I am grateful to Professor W. Whitman (University of Georgia, Athens, GA) and Professor K.F. Jarrell (Queen's University, Kingston, Ontario) for providing genomic DNA from *M. voltae* and *M. maripaludis*, and Professor E. Swiezewska (Polish Academy of Sciences, Warsaw, Poland)

for the generous gift of C55 dolichylphosphate. I would also like to thank Marcie Jaffee for assistance with cloning and purification of AglC, AglK, and AglO.

Experimental Methods

General Information

Unless otherwise noted, all solvents and reagents were obtained commercially and used without further purification. Genes were optimized for bacterial codon usage and chemically synthesized by GenScript (Piscataway, NJ). Oligonucleotides were purchased from Eurofin MWG Operon (Huntsville, AL) or Sigma Life Sciences (St. Louis, MO). Sequencing of all bacterial plasmids was conducted by the MIT CCR Biopolymers Laboratory (Cambridge, MA). Solvents were obtained from VWR (West Chester, PA) or Sigma Aldrich (St. Louis, MO). Analytical TLC was performed on silica gel 60 F254 plates (EMD Chemicals) and visualized with either UV or staining with ceric ammonium molybdate. Flash column chromatography was performed using forced flow of the indicated solvent on silica gel 60, particle size 0.043-0.063 mm (EMD Chemicals). ESI-MS was conducted on a Mariner instrument (Applied Biosystems). NMR spectra were acquired on a Varian Inova 500 MHz spectrometer with a 5 mm inverse broadband gradient probe or a Bruker 400 MHz spectrometer furnished with a 5 mm variable temperature gXH probe. Chemical shifts (δ) are reported in parts per million (ppm) and referenced to the $C\underline{D}Cl_3$ signal at 7.27 ppm for ^1H analysis and an external reference of 85% H_3PO_4 for ^{31}P NMR.

Cloning of mv155, aglC, aglK, and aglO Glycosyltransferase Genes

The *mv155, aglC, aglK,* and *aglO* genes were optimized for bacterial codon usage and chemically synthesized by GenScript. The resulting genes containing both *Bam*HI and *Xho*I restriction sites were digested and cloned into the same sites of either the pET24a(+), pGEX or pGBH vectors (Novagen) via standard molecular biology techniques. The final gene products encoded proteins with an N-terminal T7-tag and a C-terminal His_6 tag (pET24a(+)), an N-terminal GST-tag (pGEX), or an N-terminal GB1 tag and a C-terminal His_6 tag (pGBH).

Overexpression of Glycosyltransferases

The bacterial plasmids containing either the *mv155, aglC, aglK,* or *aglO* genes were transformed into *E. coli* BL21-CodonPlus(DE3) RIL or *E. coli* BL21-CodonPlus(DE3) RP competent cells (Stratagene) using the appropriate antibiotics for selection. For overexpression, 1 L of Luria-Bertani media supplemented with antibiotics was inoculated at 37 °C while shaking until an optical density (600 nm) of 0.6-0.8 was achieved. The cultures were then cooled to 16 °C and protein expression was induced by the addition of IPTG (1 mM). After 16 hrs, the cells were harvested by centrifugation (5000 x *g*) and the resultant cell pellets were stored at -80 °C until needed.

Preparation of Crude Membrane Fractions

As all of the enzymes described in this chapter are predicted to either contain a transmembrane domain or associate with the cell membrane through hydrophobic regions on the protein surface, each protein was prepared as a crude membrane fraction at some point over the course of this

study. As a general procedure, a frozen cell pellet of *E. coli* expressing the glycosyltransferase in question was resuspended in cold 50 mM HEPES, pH 8.0 / 1 mM EDTA buffer (30 mL) supplemented with Protease Inhibitor Cocktail III (Calbiochem, 30 μL). The cells were lysed by sonication, and the mixture was centrifuged at 6000 x *g* for 30 mins to remove the cellular debris. The supernatant was gently decanted and subjected to a second round of centrifugation (145,000 x *g*) for 1 hr to collect the membrane fraction. The pellet was washed with 50 mM HEPES, pH 8.0 / 1 mM EDTA buffer, centrifuged (145,000 x *g*) for 1 hr, then homogenized in cold 50 mM HEPES, pH 8.0 / 100 mM NaCl buffer (3 mL). The crude cell membrane fraction was aliquoted and stored at -80 °C for future use.

Purification of Soluble Glycosyltransferases

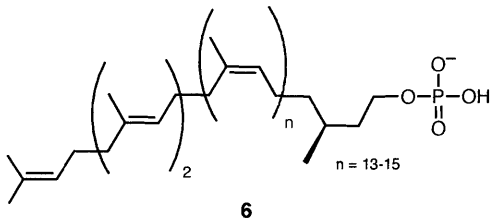
All steps were performed at 4 °C. The cell pellets from a 1 L culture were thawed and resuspended in 50 mL lysis buffer (50 mM HEPES, pH 8.0/300 mM NaCl/10 mM imidazole) with Protease Inhibitor Cocktail Set III (EDTA-free, Calbiochem) added at a 1000-fold dilution. The cells were disrupted by ultrasonication on ice, and the resulting cell lysate was then cleared of unbroken cells, cellular debris, and membranes by centrifugation (145,000 x *g*) for 65 minutes. The resulting supernatant was incubated with Ni-NTA agarose resin (Qiagen) for 2 hrs with gentle rocking and subsequently poured into a fritted PolyPrep column (BioRad) to collect the resin. The resin was washed (50 mM HEPES, pH 8.0 / 300 mM NaCl / 25 mM imidazole) and the protein eluted (50 mM HEPES, pH 8.0 / 300 mM NaCl / 250 mM imidazole). Fractions containing the desired product were combined and dialyzed (50 mM HEPES, pH 8.0 / 100 mM NaCl) to remove the imidazole and lower the salt concentration. Proteins were stored at 4 °C after the addition of glycerol to a final concentration of 25%. Protein purity was measured by

SDS-PAGE, and protein identity was confirmed by Western blot analysis using either the Anti-T7 (Novagen) or Anti-His_4 (Qiagen) antibodies. Protein concentration was determined by either the Micro BCA kit (Pierce) or UV absorbance using the appropriate extinction coefficient at 280 nm.

General Glycosyltransferase Assay

Dolichyldiphosphate *N*-acetylglucosamine (Dol-PP-GlcNAc, 5 nmol) was added to a 1.5 mL eppendorf tube and dried under a stream of nitrogen. After solubilization in DMSO (3 μL) followed by Triton X-100 (14.3%, 7 μL), 1 mM $MgCl_2$, 2.5 mM DTT, 100 mM NaCl, 50 mM HEPES, pH 8.0, and 35 μL glycosyltransferase (as a crude membrane fraction or purified protein) were combined in a final volume of 200 μL. UDP-[^{3}H]GlcNAc(3NAc)A (25 nmol, 200,000 DPM total) was added to the mixture to initiate the reaction. The reaction was incubated at room temperature for 2 hrs, after which it was quenched by the addition of $CHCl_3$:MeOH (2:1, 1.3 mL). The organic layer was extracted with pure solvent upper phase (PSUP, 3 x 300 μL) then dried under a stream of nitrogen. After resuspension in Solvable (200 μL), the mixture was diluted with Formula 989 scintillation fluid (5 mL) and analyzed by scintillation counting to determine reaction turnover.

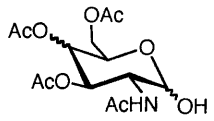
Synthesis of Dolichylphosphate



6

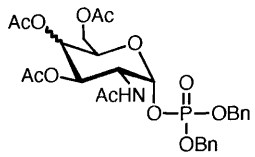
Dolichylphosphate was prepared following previously published protocols.[20,22] Dolichol (C85-C95, 50 mg in hexanes (7 μmol) was placed in a round bottom flask and dried under vacuum for 3 hrs, after which tetrazole (0.027 mmol), anhydrous bis-(2-cyanoethyl)-*N,N*-diisopropylphosphoramidite (0.011 mmol), and freshly distilled THF (1.3 mL) were added. The reaction was stirred under argon for 2 hrs, then cooled to -78 °C for 10 mins. H_2O_2 (30%, 44 μL) was then added to a final concentration of 1%, and the mixture was allowed to stir for 15 mins before warming to room temperature over the course of 15 mins. The reaction was diluted with EtOAc (5 mL), quickly washed with 1% Na_2SO_3 (5 mL) and saturated NaCl (5 mL), dried over Na_2SO_4, then concentrated under vacuum. After solubilization in NaOMe/MeOH (1%, 5 mL) to remove the cyanoethyl protecting groups, the reaction was stirred for 48 hrs at room temperature and monitored by TLC. Dolichylphosphate was purified by DE-52 anion exchange resin, eluting with a stepwise gradient of NH_4OAc in MeOH (15 mM, 25 mM, 37.5 mM, 50 mM, 100 mM) to yield roughly 25 mg (~50%) of the desired product **6**. ^1H and ^{31}P- NMR data matched previously reported values.[23]

Synthesis of Dolichol-Linked Monosaccharides



9 GlcNAc
10 GalNAc

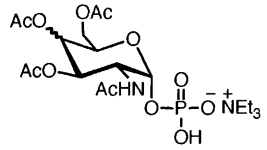
*a. 3,4,6-Tri-O-acetyl-2-deoxy-2-acetamido-α,β-D-glucose (**9**) and galactose (**10**):* To a solution of D-glucose (**7**) or D-galactose (**8**, 1.0 g, 4.52 mmol) in pyridine (25 mL), acetic anhydride (4.26 mL, 45.2 mmol) was added. The reaction was stirred overnight at room temperature, after which the solvent was removed under vacuum and the crude reaction mixture was recrystallized from EtOH. In order to deprotect the anomeric hydroxyl group, a solution of the tetracetate (2.5 g, 6.42 mmol) and dimethylamine (16.0 mL, 32.1 mmol) was stirred in CH_3CN (100 mL) at room temperature for 24 hrs. The mixture was diluted with EtOAc (40 mL) and washed with H_2O (40 mL) and NaCl (40 mL), and then dried (Na_2SO_4). The product was purified over silica gel with 9:1 CH_2Cl_2:MeOH to give the desired triacetylated sugars **9** and **10** as a mixture of α and β anomers. ^1H NMR and ESI-MS data match published values.[24]



11 GlcNAc
12 GalNAc

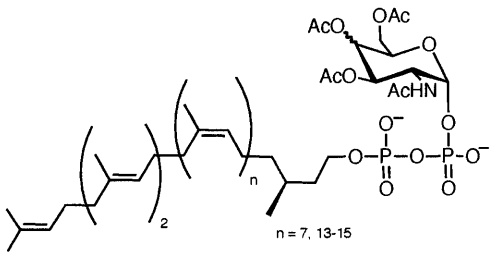
*b. Dibenzyl 3,4,6-tri-O-acetyl-2-deoxy-2-acetamido-α-D-glucose phosphate (**11**) and -galactose phosphate (**12**):* Dibenzyl *N,N*-diethylphosphoramidite (0.86 g, 7.2 mmol) was added to a stirring solution of the triacetylated glucose (**9**, 1.0 g, 2.9 mmol) or galactose (**10**, 1.0 g, 2.9

mmol) and 1,2,4-triazole (0.8 g, 11.5 mmol) in anhydrous CH_2Cl_2 under a nitrogen atmosphere at room temperature. The mixture was allowed to stir for 2 hrs before being diluted with Et_2O. The mixture was then washed with cold $NaHCO_3$, NaCl, and H_2O, dried (Na_2SO_4), and concentrated under vacuum. To oxidize the phosphite, the reaction mixture was resuspended in THF (40 mL) and cooled to -78 °C. H_2O_2 (30%, 8 mL) was added to the stirring solution in a dropwise manner, after which the reaction was chilled for an additional 15 mins, then allowed to warm to room temperature over the course of 90 mins. The reaction mixture was diluted with Et_2O and washed with cold $NaHCO_3$, NaCl, and H_2O in succession, dried (Na_2SO_4), and concentrated under vacuum as described above. The crude product was purified over silica gel with 1:4 EtOAc:hexanes to give the desired phosphate (**11** or **12**). ^1H and ^{31}P NMR and ESI-MS data match published values.[24]



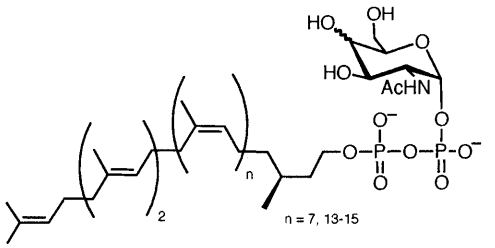
13 GlcNAc
14 GalNAc

c. *3,4,6-tri-O-acetyl-2-deoxy-2-acetamido-α-D-glucose phosphate (**13**) and -galactose phosphate (**14**):* The dibenzyl phosphates **11** and **12** (1.0 g, 1.8 mmol) were subjected to hydrogenation over 5% $Pd(OH)_2/C$ (150 mg) in EtOH (30 mL) and 10% $NaHCO_3$ (20 mL) for 12 hrs at room temperature. The mixture was filtered, and the filtrate concentrated under vacuum. The residue was treated with a solution of 1:1 MeOH:H_2O in 10% Et3N, then concentrated to give the phosphates **13** and **14**. ^1H and ^{31}P NMR and ESI-MS data match published values.[24]



15 n = 7, GlcNAc **17** n = 7, GalNAc
16 n = 13-15, GlcNAc **18** n = 13-15, GalNAc

d. General procedure for coupling of sugar phosphates with dolichylphosphates: To a solution of the sugar phosphate (12.0 mg, 0.02 mmol) in dry DMF (1 mL) at room temperature was added 1,1'carbonyldiimidazole (CDI, 0.2 mmol). After 6 hrs, MeOH was added to the reaction in a dropwise manner to quench excess CDI, and the mixture was incubated for 30 mins. Dolichylphosphate (either C55 or C85-C95, 0.014 mmol) in CH_2Cl_2 (1 mL) was then added, and the reaction was allowed to stir for 8 days at room temperature and monitored by TLC. Upon completion, the crude mixture was concentrated under vacuum and purified by silica gel chromatography (85:14:1 $CHCl_3$:MeOH:2M NH_4OH) to give the pure triacetylated dolichyl-linked monosaccharides **15**, **16**, **17**, and **18**. ^{31}P NMR data match published values.[23]



1 n = 7, GlcNAc **3** n = 7, GalNAc
2 n = 13-15, GlcNAc **4** n = 13-15, GalNAc

e. General procedure for deprotection of acetylated dolichyl-linked monosaccharides: The protected dolichyl-linked monosaccharides were dissolved in MeOH (1 mL) and a solution of

25% NaOMe/MeOH (40 μL) was added dropwise to a final concentration of 1% NaOMe. The solution was stirred at room temperature for 30 mins, then neutralized with Dowex 50WX8-200 ion exchange resin (pyridinium form) and concentrated under vacuum to yield the desired dolichyl-linked monosaccharides **1**, **2**, **3**, and **4**. ^{31}P NMR values (162 MHz, 5:1 $CDCl_3$:CD_3OD) are as follows:

Dolichyldiphosphate(C55)-N-acetyl-α-D-glucosamine (**1**): δ -5.8, -8.6

Dolichyldiphosphate(C85-C95)-N-acetyl-α-D-glucosamine (**2**): δ -17.7, -18.9

Dolichyldiphosphate(C55)-N-acetyl-α-D-galactosamine (**3**): δ -5.3, -8.1

Dolichyldiphosphate(C85-C95)-N-acetyl-α-D-galactosamine (**4**): δ -16.3, -17.8

References

1. Varki, A.; Cummings, R.D.; Freeze, H.H.; Stanley, P.; Bertozzi, C.R.; Hart, G.W.; Etzler, M.E. *Essentials of Glycobiology*; 2nd Ed.; Cold Spring Harbor Laboratory Press: New York, 2009.
2. Lairson, L.L.; Henrissat, B.; Davies, G.J.; Withers, S.G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521-555.
3. Coutinho, P.M.; Deleury, E.; Davies, G.J.; Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **2003**, *328*.
4. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active enZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233-D238.
5. Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koča, J.; Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **2006**, *16*, 29R-37R.
6. Vrielink, A.; Rüger, W.; Driessen, H.P.C.; Freemont, P.S. Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J.* **1994**, *13*, 3413-3422.
7. Chiu, C.P.; Watts, A.G.; Lairson, L.L.; Gilbert, M.; Lim, D.; Wakarchuk, W.W.; Withers, S.G.; Strynadka, N.C. Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. *Nat. Struct. Mol. Biol.* **2004**, *11*, 163-170.
8. Lairson, L.L.; Withers, S.G. Mechanistic analogies amongst carbohydrate modifying enzymes. *Chem. Commun. (Camb)* **2004**, *20*.
9. Unligil, U.M.; Rini, J.M. Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **2000**, *10*, 510-517.
10. Gao, X.-D.; Moriyama, S.; Miura, N.; Dean, N.; Nishimura, S.-I. Interactions between the C-Termini of Alg13 and Alg14 mediates formation of the active UDP-*N*-acetylglucosamine transferase complex. *J. Biol. Chem.* **2008**, *283*, 32534-32541.
11. Wang, X.; Weldeghorghis, T.; Zhang, G.; Imperiali, B.; Prestegard, J.H. Solution structure of Alg13: the sugar donor subunit of a yeast *N*-acetylglucosamine transferase. *Structure* **2008**, *16*, 965-975.
12. Glover, K.J.; Weerapana, E.; Imperiali, B. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14255-14259.
13. Chaban, B.; Logan, S.M.; Kelly, J.F.; Jarrell, K.F. AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the *N*-Glycan in *Methanococcus voltae*. *J. Bacteriol.* **2009**, *191*, 187-195.
14. VanDyke, D.J.; Wu, J.; Logan, S.M.; Kelly, J.F.; Mizuno, S.; Aizawa, S.-I.; Jarrell, K.F. Identification of genes involves in the assembly and attachment of a novel flagellin *N*-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripaludis*. *Mol. Microbiol.* **2009**, *72*, 633-644.
15. Chaban, B.; Voisin, S.; Kelly, J.; Logan, S.M.; Jarrell, K.F. Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae N*-linked glycans: insight into *N*-linked glycosylation pathways in Archaea. *Mol. Microbiol.* **2006**, *61*, 259-268.

16. Yurist-Doutsch, S.; Chaban, B.; VanDyke, D.J.; Jarrell, K.F.; Eichler, J. Sweet to the extreme: protein glycosylation in archaea. *Mol. Microbiol.* **2008**, *68*, 1079-1084.
17. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes *J. Mol. Biol.* **2001**, *305*, 567-580.
18. Lechner, J.; Wieland, F.; Sumper, M. Biosynthesis of sulfated saccharides N-glycosidically linked to the protein via glucose. Purification and identification of sulfated dolichyl monophosphoryl tetrasaccharides from halobacteria. *J. Biol. Chem.* **1985**, *260*, 860-866.
19. Larkin, A.; Imperiali, B. Unpublished work, 2010.
20. Ye, X.-Y.; Lo, M.-C.; Brunner, L.; Walker, D.; Kahne, D.; Walker, S. Better substrates for bacterial transglycosylases. *J. Am. Chem. Soc.* **2001**, *123*, 3155-3156.
21. Branch, C.L.; Burton, G.; Moss, S.F. An expedient synthesis of allylic polyprenol phosphates. *Synth. Commun.* **1999**, *29*, 2639-2644.
22. Weerapana, E.; Glover, K.J.; Chen, M.M.; Imperiali, B. Investigating bacterial *N*-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **2005**, *127*, 13766-13767.
23. Tai, V.W.F.; Imperiali, B. Substrate specificity of the glycosyl donor for oligosaccharyl transferase. *J. Org. Chem.* **2001**, *66*, 6217-6228.
24. Sim, M.M.; Kondo, H.; Wong, C.-H. Synthesis and use of glycosyl phosphites: an effective route to glycosyl phosphates, sugar nucleotides, and glycosides. *J. Am. Chem. Soc.* **1993**, *115*, 2260-2267.

Chapter 6 Efforts Towards the Inhibition of PglB, the Oligosaccharyl Transferase from the Bacterial Pathogen *Campylobacter jejuni*

Introduction

Asparagine-linked glycosylation is a posttranslational protein modification that occurs in all domains of life.[1-3] This process entails the stepwise assembly of an oligosaccharide onto a polyisoprenyl-linked carrier, followed by the *en bloc* transfer of the glycan to asparagine residues of nascent or fully folded proteins (Figure 6-1). In eukaryotes, this process is carried out by the multimeric enzyme complex oligosaccharyl transferase (OT), localized in the lumen of the endoplasmic reticulum. As outlined in Chapter 2, OT transfers the tetradecasaccharide ($GlcNAc_2Man_9Glc_3$) from dolichyl-pyrophosphate to nascent proteins, in which the Asn residue is found within an Asn-Xaa-Ser/Thr sequon (where Xaa is any amino acid except proline) and aligned in an Asx turn.[4,5]

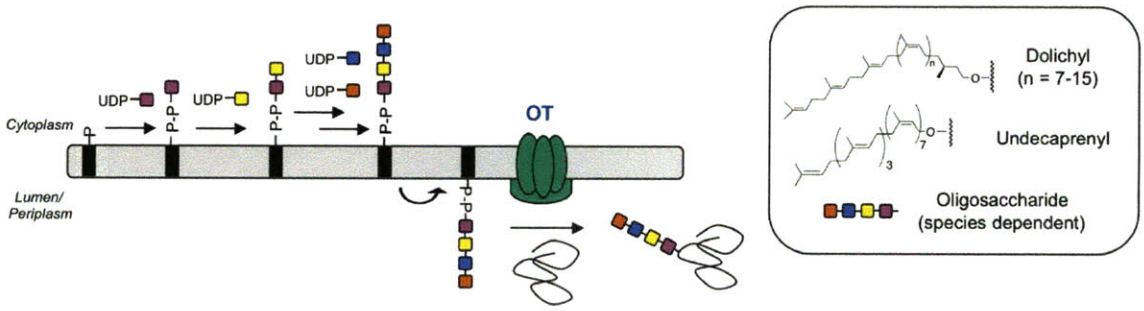


Figure 6-1: General pathway of *N*-linked glycosylation.

Although OT has been the subject of intense investigation over several decades, a detailed understanding of the reaction mechanism of this complex has remained elusive. One of the most intriguing questions concerning this mechanism involves the nucleophilicity of the asparagine side chain; it is still unclear how the amide nitrogen, a stable and fairly unreactive moiety, is activated for catalysis within the enzyme active site. There have been several proposals put forth to describe this decisive step, as summarized in Figure 6-2. The first

proposal, described by Marshall (Figure 6-2A), suggests that deprotonation of the amide nitrogen is enhanced by a hydrogen bond between the Asn side chain carbonyl and the hydroxyl group of the required Ser/Thr residue.[6] A second model by Bause and coworkers states that the amide nitrogen behaves as a hydrogen bond donor, and is activated through deprotonation by the Ser/Thr hydroxyl oxygen atom (Figure 6-2B).[7] In contrast to these two models, the proposal put forth by Imperiali highlights the importance of the peptide backbone conformation in the reaction mechanism, specifically by describing the critical role of the Asx turn (Figure 6-2C).[8] This motif is believed to provide a series of key hydrogen bonding interactions that aid in deprotonation of the amide nitrogen to form an imidol intermediate, which then reacts with the dolichyl-linked oligosaccharide to catalyze glycosyl transfer.

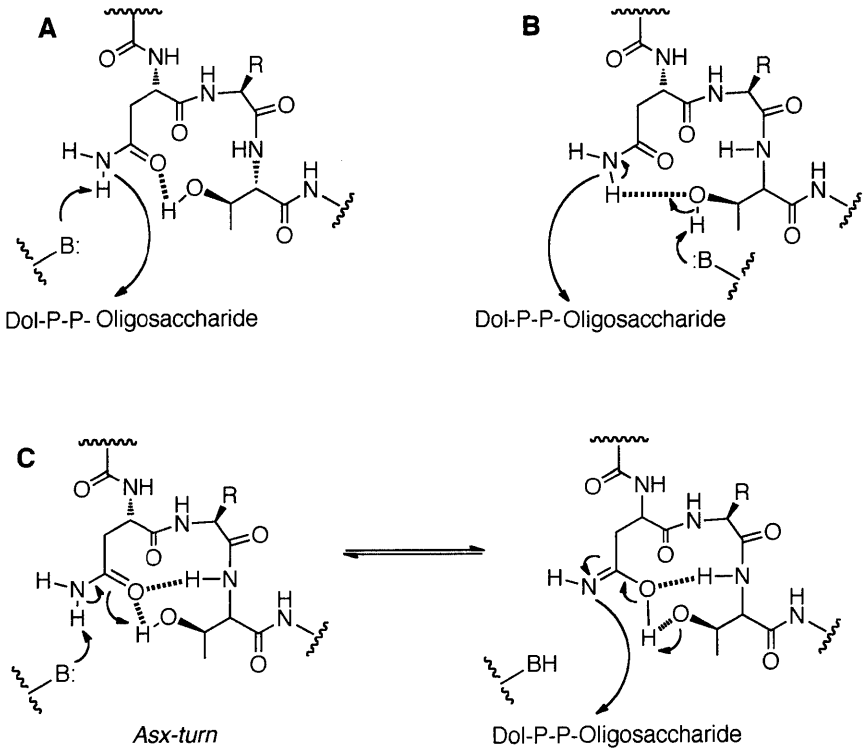
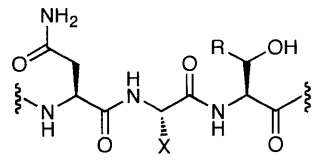


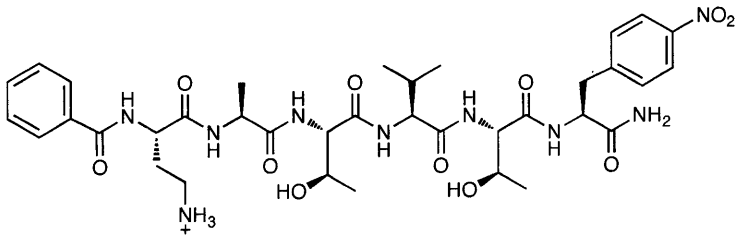
Figure 6-2: Three mechanistic proposals for the activation of the amide side chain nitrogen of asparagine by OT. (A) Marshall; (B) Bause; (C) Imperiali.

A powerful method for studying enzyme mechanism is through the use of substrate-based inhibitors. This classical approach involves the design and synthesis of molecules that mimic an intermediate along the reaction pathway, thus competing with the natural substrate to bind to the enzyme and resulting in a reduced overall reaction rate. Previous work in the Imperiali group has yielded a number of peptide-based inhibitors that exhibit low nanomolar affinity for OT. The structures of these compounds are based upon the Asn-Xaa-Ser/Thr tripeptide glycosylation sequence (**1**) derived from the native protein substrates (Figure 6-3). Early studies determined that the use of isosteres in the place of the Asn residue resulted in competitive inhibition of OT; these isosteres are substituted for the asparagine side chain yet prevent glycan transfer to the peptide.[9] It was determined that the Bz-Dab-Ala-Thr-Val-Thr-Nph-NH_2 peptide (**2**), in which the Asn residue is replaced by 2,4-diaminobutanoic acid (Dab), was a potent inhibitor of the OT complex in *Saccharomyces cerevisiae* in vitro (K_i = 62 nm).[10,11] In addition, constraining the peptide into an Asx-turn resulted in increased affinity for the enzyme, as in the case of **3**, which highlights the importance of the Asx turn for enzyme catalysis. Other efforts to develop OT inhibitors led to the generation of the glycopeptide mimetic (**4**),[12] as well as the dolichyl-linked disaccharide (**5**), in which the 2-acetamido group of the first GlcNAc residue is replaced by a fluorine atom.[13] Studies of OT activity using **5** indicated that it was not accepted as a glycosyl donor, providing the first evidence that the 2-acetamido group on the proximal sugar is necessary for substrate turnover; this requirement has since been confirmed for OT homologs in prokaryotes.[14] However, despite the wealth of information acquired using these inhibitors, difficulties in obtaining highly pure OT has complicated efforts to gain a detailed biochemical characterization of the enzyme mechanism.



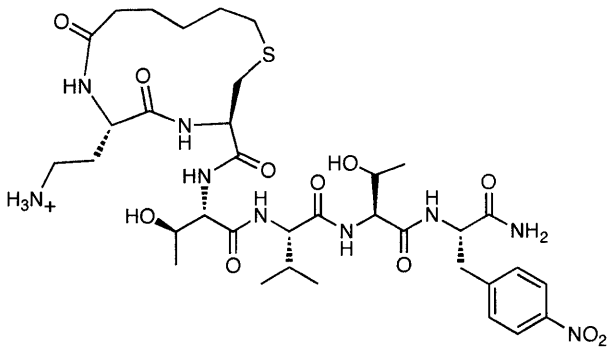
1

Asn-Xaa-Ser/Thr R = H (Ser), CH_3 (Thr)
Native Glycosylation Sequence



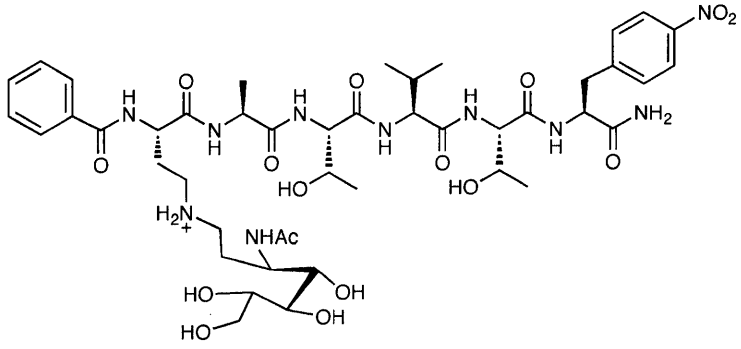
2

Bz-Dab-Ala-Thr-Val-Thr-Nph-NH_2
K_i = 62 nM



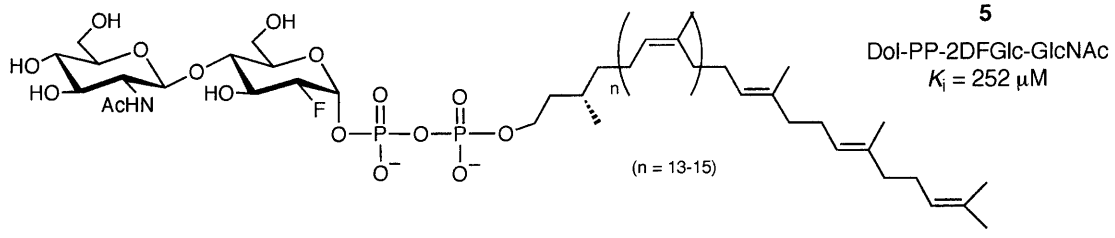
3

c[Hex-Dab-Cys-]Thr-Val-Thr-Nph-NH_2
K_i = 37 nM



4

Bz-Dab(GlcNAc)-Ala-Thr-Val-Thr-Nph-NH_2
K_i = 41 nM



5

Dol-PP-2DFGlc-GlcNAc
K_i = 252 μM

Figure 6-3: Previous OT inhibitors developed by the Imperiali laboratory.

In 1999, a general system of *N*-linked glycosylation was identified in the pathogenic, Gram-negative bacterium *Campylobacter jejuni*.[15] *C. jejuni* is a human gut mucosal pathogen that is a leading cause of bacterial gastroenteritis,[16,17] and the presence of *N*-linked glycoproteins has been found to play a key role in virulence.[18] As described in Chapter 1, the *C. jejuni* oligosaccharyl transferase PglB catalyzes the transfer of the heptasaccharide ($BacGalNAc_5Glc$) from an undecaprenyl-pyrophosphate donor onto the amide side chain nitrogen of acceptor proteins.[19] Unlike the multimeric OT complex found in eukaryotes, PglB is a single polypeptide composed of two domains, a large N-terminal region made up of roughly 9-13 transmembrane helices, and a soluble C-terminal domain that contains the canonical WWDXGX motif, which is conserved in the catalytic subunit of the eukaryotic OT, Stt3p, as well as all putative OT homologs.[20] While PglB recognizes the N-X-S/T sequon for glycosylation of the target polypeptide, similar to OT, it also has the additional requirement of the negatively charged residues D/E at the -2 position. Despite these subtle differences, it is proposed that the mechanisms of both OT and PglB are highly similar, and thus the study of PglB provides a simpler system that is more suitable for in depth biochemical characterization.

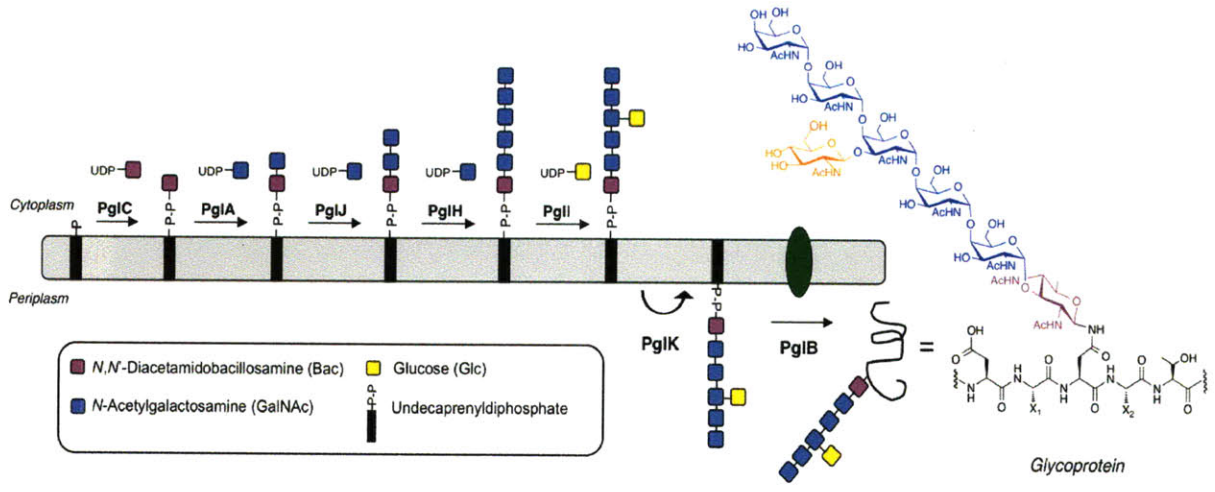


Figure 6-4: *N*-linked glycosylation in *C. jejuni.*

In this chapter, efforts to gain insight into the enzymatic mechanism of the bacterial oligosaccharyl transferase PglB are described. A major focus to this end involved the synthesis and testing of over forty potential peptide-based inhibitors, many of which incorporated isosteres of asparagine. However, despite trying a wide variety of reaction conditions, no inhibition of PglB was observed using these peptides. In an alternative approach, the use of a panel of small molecule enzyme alkylating agents is described in an attempt to identify key amino acids involved in the mechanism of PglB. This study revealed that diethylpyrocarbonate (DEPC), a histidine-specific alkylating agent, was able to selectively inhibit the enzyme in a time and concentration dependent manner. A series of PglB point mutations was then prepared, in which each of the five histidines found in the protein sequence was sequentially mutated to alanine, in an effort to identify the critical histidine residue in question. This work summarizes efforts towards the search for PglB inhibitors, which we envision would be a powerful tool to gain insight into the mechanism of the oligosaccharide transferase as well as a starting point in the design of novel therapeutics against the human mucosal pathogen *C. jejuni*.

Results and Discussion

Peptide-based Inhibitors of PglB

In order to study the active site requirements of PglB, a series of peptide-based inhibitors was designed, synthesized, and tested against the enzyme. Each of these peptides is based on the Asp-Xaa_1-Asn-Xaa_2-Ser/Thr-Nph scaffold, where Xaa_1 is Phe/Gln, and Xaa_2 is Ala/Val (Figure 6-5). Previous work in the Imperiali group found this peptide sequence to be optimal for glycosylation by PglB.[21] These peptides are acetylated at the N-terminus and include a C-

terminal *p*-nitrophenylalanine residue (Nph) to aid in quantification. All peptides were synthesized on PAL-PEG-PS resin using standard Fmoc-solid phase peptide techniques, purified by RP-HPLC, and characterized by mass spectrometry.

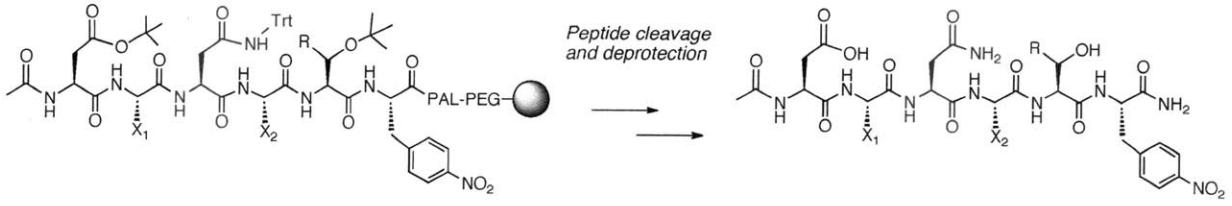


Figure 6-5: General scheme for synthesis of PglB peptide inhibitors on PAL-PEG-PS resin. Residue Xaa_1 is Phe or Gln and Xaa_2 is Ala or Val, and R is either -H (Ser) or $-CH_3$ (Thr).

A standard assay was used to screen PglB for activity, based upon that originally developed for the yeast OT (Figure 6-6).[19,22] It involves the incubation of PglB with the radiolabeled disaccharide donor Und-PP-Bac-[^{3}H]GalNAc and other reaction components, including detergent and $MnCl_2$. The assay is initiated with the addition of the peptide substrate, and aliquots are removed at various time points and quenched by aqueous/organic mixture, which serves to partition the glycosylated peptide into the aqueous phase and the isoprenyl-starting material into the organic phase. Glycosylation of the peptide substrate by PglB results in a net transfer of radioactive counts from the organic phase into the aqueous phase as measured by scintillation counting. In order to screen potential PglB inhibitors for activity, PglB first is incubated with the inhibitor peptide (50-200 μM), the radiolabeled glycan donor, and other reaction components at room temperature for 10-30 minutes. The glycosylation reaction is then initiated by the addition of the peptide substrate (10-50 μM), and the assay is carried out as described above. Comparison of reaction rates of PglB incubated with and without the potential inhibitor is then performed in order to determine whether inhibition has taken place.

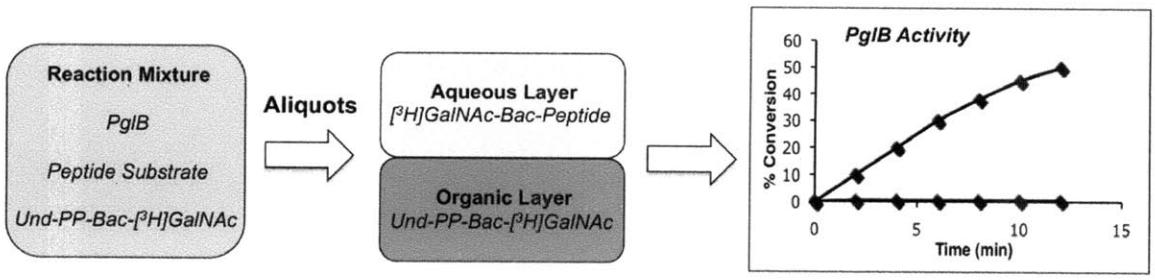


Figure 6-6: Assay for PglB activity. PglB is incubated with the peptide substrate, the radiolabeled glycan donor, and other reaction components and aliquots are removed over time and subjected to phase extraction. Glycosylation results in the transfer of radioactive counts from the organic to the aqueous phase. In the case of PglB inhibition, fewer counts are transferred to the aqueous phase corresponding to less overall activity.

a. Asparagine isosteres

The majority of peptides synthesized and tested for inhibition of PglB contain isosteres of asparagine, in which the peptide scaffold remains unchanged but the asparagine side chain amide is altered in some way (Figure 6-7). By modifying only the asparagine amide, the peptide can presumably maintain extended binding contacts with the enzyme, but can no longer serve as a glycan acceptor. Each of the isosteres included in this study was designed to test a particular feature of the requirements for glycosylation; for example, the length of the side chain is varied through incorporation of glutamine (**6**), while the tolerance for charge is explored in **9** and **11**. In the case of peptides **12-14**, the amide bond is preserved, but the asparagine is extended with nonpolar (**12**) or polar (**13**) and nucleophilic (**14**) moieties. For peptides **15-23**, moieties that mimic the stereoelectronic properties of asparagine were used in the place of the side chain amide. The synthesis of these peptides was accomplished through one of three ways: a) incorporation of a commercially available amino acid during peptide synthesis (**6**, **7**, **9, 10**, **21-23**); b) synthesis of the isosteric amino acid, followed by its inclusion into the peptide framework (**11**, **15-17, 19, 20**); or c) introduction of the isostere after completion of the peptide

backbone (**8**, **12-14, 18**). The following sections summarize the design and synthesis of these isosteric peptides.

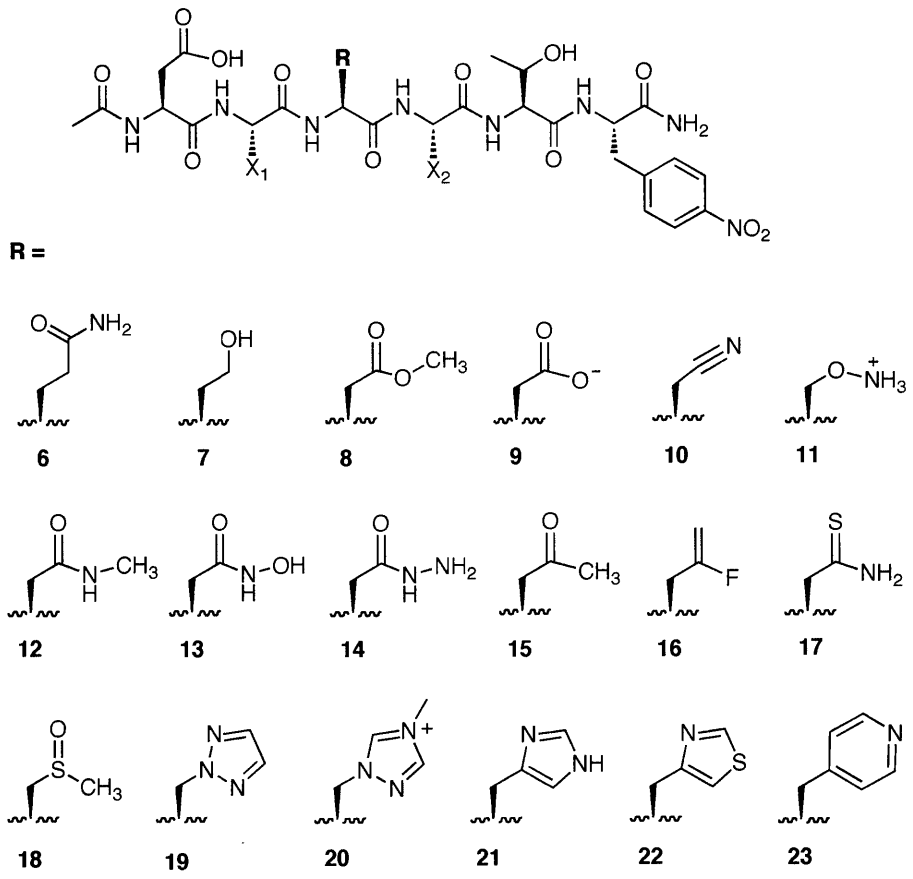


Figure 6-7: PglB peptides containing asparagine isosteres outlined in this study.

L-Oxonorvaline (15)

Several classes of inhibitory molecules, such as those directed against proteases, often contain peptidyl ketone and aldehyde functionalities.[23] These molecules utilize the electrophilic carbonyl center as a trap for nucleophiles located within the enzyme active site. In order to produce a similar trap for PglB, a methyl ketone moiety was desired as a side chain replacement of the asparagine amide. It was postulated that if the catalytic mechanism of PglB involves deprotonation of the amide as an initial step, one might envision the formation of the reactive

enolate that may be useful for mechanistic studies. A synthesis of the methyl ketone amino acid was modified from previously published procedures (Figure 6-8),[24,25] starting with the readily available L-Boc-Asp-OtBu amino acid (**24**). Conversion to the primary alcohol (**25**) was achieved via the mixed anhydride, followed by reduction with $NaBH_4$ in H_2O. The Swern oxidation was then employed to give the aldehyde (**26**) in good yield, followed by alkylation using diazomethane to afford the ketone (**27**). After deprotection of the amine and ion exchange with Dowex (Cl^-) resin, the Fmoc protecting group was installed to yield the desired product (**29**), which was incorprated into the PglB peptide scaffold using standard Fmoc peptide chemistry as described. No inhibition was observed when this peptide was tested with PglB.

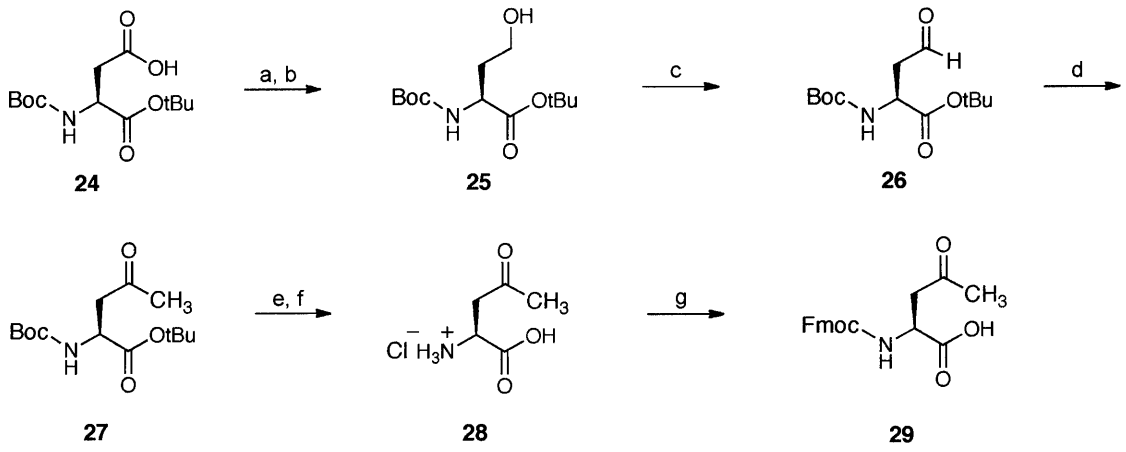
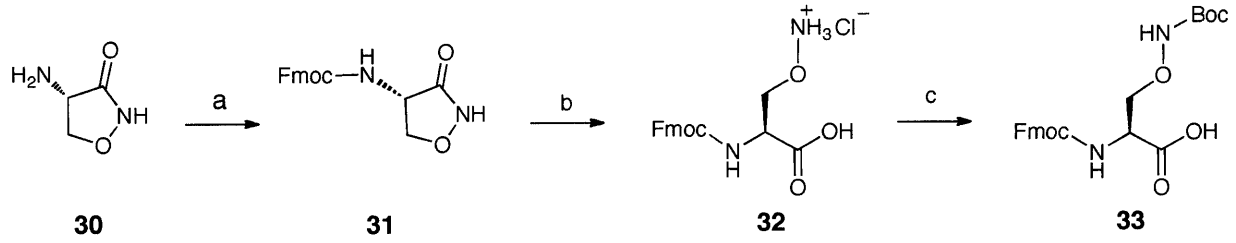


Figure 6-8: Synthesis of the methyl ketone amino acid, Fmoc-L-oxonorvaline-OH (**29**). Reagents and conditions: (a) CH_3CH_2OCOCl, NEt_3, THF, -5 °C; (b) $NaBH_4$, H_2O, 10 °C; (c) ClCOCOCl, DMSO, DIPEA; (d) CH_2N_2, CH_2Cl_2; (e) TFA, CH_2Cl_2; (f) Dowex (Cl^-), MeOH; (g) Fmoc-OSu, $NaHCO_3$, 1,4-dioxane/MeOH.

L-Aminoserine (11)

Another asparagine isostere of interest was L-aminoserine (Ams, **6**), which contains a nucleophilic oxyamine functionality in the place of the side chain amide. The oxyamine retains the hydrogen bond donor capabilites of the amide nitrogen, but is compromised in its ability to

accept a hydrogen bond compared with amide carbonyl, which may affect formation of the required Asx turn. In addition to its role as an asparagine isostere, it was envisioned that this compound could serve as a handle for the chemoselective ligation of carbohydrates and other aldehyde-bearing compounds of interest.[12] The Ams amino acid was previously synthesized in the Imperiali group, and peptides incorporating this amino acid residue exhibited a K_i value of 34 μM against the yeast OT.[26] The aminoserine amino acid was synthesized according to the scheme outlined in Figure 6-9, and involves the Fmoc-protection of the N-terminal amine of L-cycloserine (**30**), followed by acid-promoted ring opening and protection of the oxime nitrogen with Boc_2O to yield the final product (**33**).[27] The amino acid was then incorporated into the PglB peptide sequence for testing against the enzyme; however, no inhibition of the enzyme was observed.



30 **31** **32** **33**

Figure 6-9: Synthesis of Fmoc-L-aminoserine(Boc)-OH (**32**). Reagents and conditions: (a) Fmoc-OSu, $NaHCO_3$, 1,4-dioxane/H_2O; (b) 2M HCl, H_2O/THF; (c) Boc_2O, DIPEA, DMF.

L-Thioasparagine (17)

Another isostere explored for potential inhibition of PglB was the thioasparagine (**17**). Thioasparagine is a very close isostere of asparagine, wherein the carbonyl is replaced with the bulky, nucleophilic thionyl; as expected, the pK_a values of these two side chains are predicted to be quite similar (15.1 for acetamide and 13.4 for thioacetamide).[28,29] Although this amino acid

was previously synthesized by the Imperiali group and found to be accepted as a substrate for OT (K_m = 250 μM), the procedure involved the conversion of a cyano group to the thioamide using a harsh combination of H_2S and NH_3 gases. In an effort to avoid this potentially dangerous reaction, we instead sought to synthesize the desired amino acid as summarized in Figure 6-10. Protection of the C-terminal carboxylate of Fmoc-Asn-OH (**34**) as the paramethoxybenzyl ester (PMB, **35**) was followed by thiolation of the carbonyl with P_2S_5 and sonication. Removal of the PMB group and subsequent protection of the thioamide side chain with a xanthenyl moiety afforded the desired amino acid (**38**).

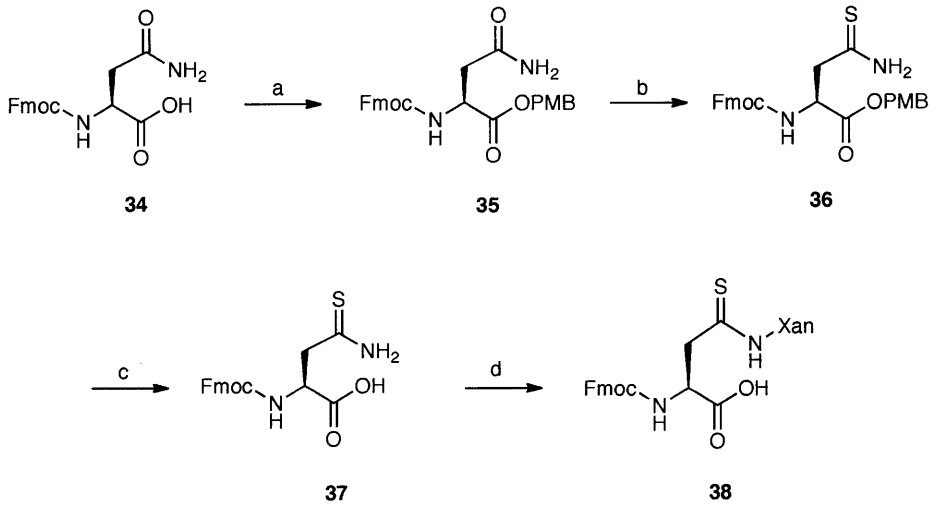


Figure 6-10: Synthesis of Fmoc-thioasparagine(Xan)-OH (**37**). Reagents and conditions: (a) $ClCH_2PhOMe$, DIPEA, NaI, DMA, N_2, rt; (b) P_2S_5, THF, Ar, sonication; (c) thioanisole, EDT, TFA; (d) 9-hydroxanthene, HOAc, 85 °C.

Despite several attempts, efforts to incorporate this amino acid into the peptide framework proved unsuccessful. A variety of peptide coupling agents, including PyBOP, PyAOP, HATU, DIC, HBTU, and HOAt were utilized to activate the carboxylate, but the coupling of this amino acid to the peptide scaffold was not achieved. It was hypothesized that perhaps the bulkiness of both the xanthydrol and Fmoc protecting groups were preventing

activation of the carboxyl terminus through steric hindrance. An alternate strategy for generating the desired peptide involved the synthesis of an asparagine-containing peptide, followed by selective conversion of the side chain amide to the thioamide using Lawesson's reagent (Figure 6-11). HPLC analysis of the free peptide indicated only minimal conversion of the carbonyl to the thionyl (**17**); however, just enough material was achieved in order to test PglB for inhibition. Unfortunately, as in the case with other asparagine isosteres described in this section, this peptide was neither a substrate nor an inhibitor of the PglB reaction.

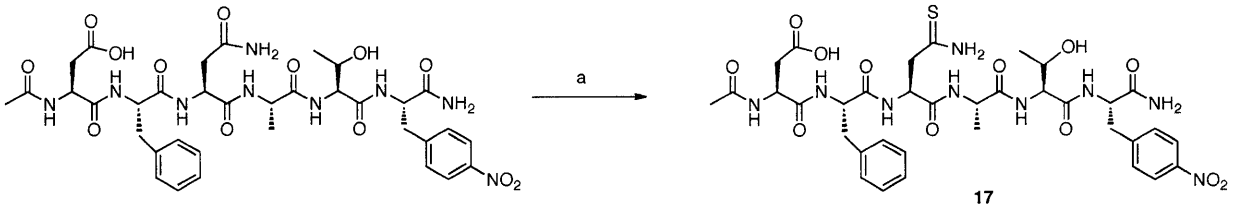


Figure 6-11: Synthesis of the thioasparagine-containing peptide. Reagents and conditions: (a) Lawesson's reagent, THF:DMSO (1:1), sonication, 8 hrs.

*L-Methyl Sulfoxide (**18**)*

An asparagine isostere bearing a methyl sulfoxide side chain (**18**) was also prepared. This amino acid differs from asparagine in several ways, including the larger size of the sulfur atom compared to carbon and the weaker electron donating capability of the sulfoxide oxygen, and may also serve as an electrophilic trap for active site nucleophiles. The decreased acidity of a methyl sulfoxide ($pK_a \sim 33$) compared with a carboxamide ($pK_a = 15.1$ for acetamide) suggests that this peptide would not be glycosylated, particularly if the mechanism of glycosylation involves deprotonation as a first step; therefore, it could be a useful competitive inhibitor. The synthesis of **18** began with the incorporation of L-(S-methyl)-cysteine into the peptide scaffold using standard peptide chemistry techniques. This peptide (**39**) was then

released from the solid support and deprotected with acid treatment, purified, and oxidized to the sulfoxide (**18**) with $NaIO_4$ (Figure 6-12).[30,31] HPLC purification of the reaction mixture allowed for resolution of both diastereomers, which were formed in an approximate 1:1 ratio. Both the methyl cysteine and sulfoxide peptides were then tested for inhibition activity against PglB.

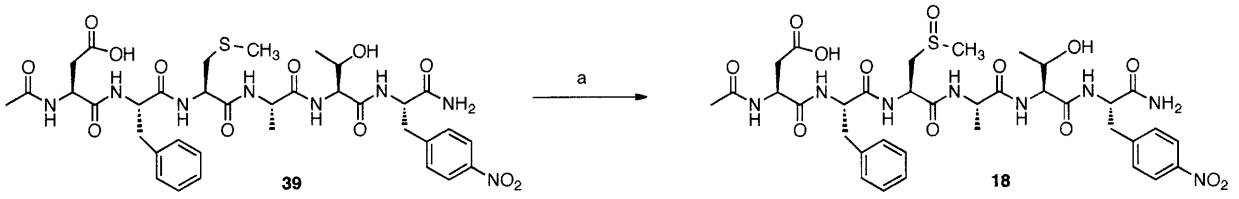


Figure 6-12: Synthesis of the methyl sulfoxide-containing peptide (**18**). Reagents and conditions: (a) $NaIO_4$, $MeOH/H_2O$, 0 °C (1 hr), then 25 °C (24 hrs).

Extended Asparagines: L-MethylAsn (12), L-HydroxyAsn (13), L-HydrazidoAsn (14)

A set of three "extended asparagine" peptides was explored as a means to determine whether PglB could tolerate a slightly longer residue in the place of the native amide side chain. These peptides were generated by first synthesizing a peptide in which an allyl-protected aspartate residue was substituted for the asparagine. After selective deprotection of the carboxylic acid side chain with $Pd(PPh_3)_4$ in the presence of phenylsilane, the acid was coupled with either methylamine, hydroxylamine, or *t*-butylcarbazate and PyAOP/HOAt (Figure 6-13). The peptides were then globally deprotected and cleaved from the resin using a standard cleavage cocktail to yield the isosteres **12**, **13**, and **14**, respectively. Previous work in the Imperiali group indicated that these peptides were modest inhibitors of OT; for example, the hydrazine isostere exhibited a K_i of 2.5 μM.[12] Upon testing these peptides with PglB, we observed that the hydroxyasparagine (**13**) was in fact a poor substrate, with a K_m of > 100 μM.

However, both of the other extended asparagine peptides (**12** and **14**) exhibited neither substrate nor inhibitor activity.

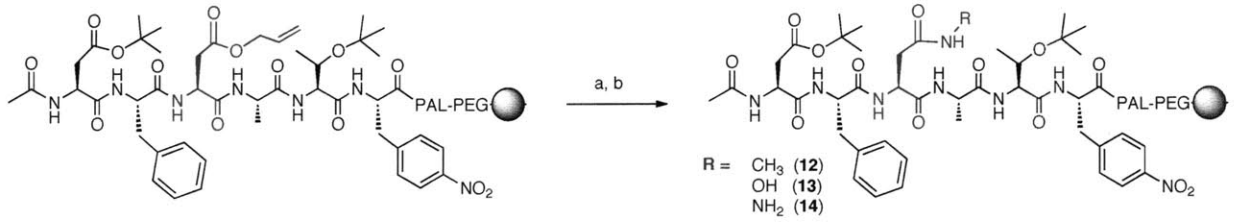


Figure 6-13: Synthetic strategy for extended asparagines **12-14**. Reagents and conditions: (a) $Pd(PPh_3)_4$, phenylsilane, CH_2Cl_2 (3x); (b) NH_2CH_3 (**12**), NH_2OH (**13**), or $NH_2NHCOOtBu$ (**14**), PyAOP, HOAt, DMF.

*Other isosteres: Fluoropropenylalanine (**16**), Histidine Analogs (**19-23**)*

The final set of isosteric peptides involves the fluoropropenylalanine (**16**), histidine (**21**), and histidine analogs (**19, 20, 22, 23**), all of which present variations in the stereoelectronic profile of the side chain compared with asparagine. All of these peptides were produced through the incorporation of the relevant amino acids into the peptide backbone as previously described. The fluoropropenyl amino acid was provided by Günter Haufe,[32] while L-2-(1,2,3-triazolyl)alanine and L-(4-methyl-1,2,4-triazolium)alanine were both obtained from the Haines group (Wellesley College). Histidine, thiazolylalanine, and 4-pyridinylalanine were also installed in the place of asparagine to give **21**, **22**, and **23** respectively. Unfortunately, none of these peptides displayed activity as either PglB substrates or inhibitors.

b. Bisubstrate Inhibitors: Aryl Asparagines/Neoglycopeptides

Another approach to designing PglB inhibitors involves use of the bisubstrate inhibitor, a compound that combines structural features of two separate enzyme substrates. These types of inhibitors are widely utilized in the pharmaceutical industry in order to achieve increased binding affinity of a small molecule for the target enzyme. Inspiration for applying this method to the design of PglB inhibitors was derived from tunicamycin, which to date is the only commercially available inhibitor of *N*-linked glycosylation. Tunicamycin, first isolated from the soil bacterium *Streptomyces lysosuperificus* in 1971,[33] is a potent inhibitor of the first committed step in the *N*-linked glycosylation pathway in eukaryotes carried out by the GlcNAc-1-P phosphoglycosyltransferase (Alg7 in *S. cerevisiae*).[34] As depicted in Figure 6-14, the structure of tunicamycin incorporates features of both the UDP-GlcNAc and Dol-PP-GlcNAc substrates of this enzyme.

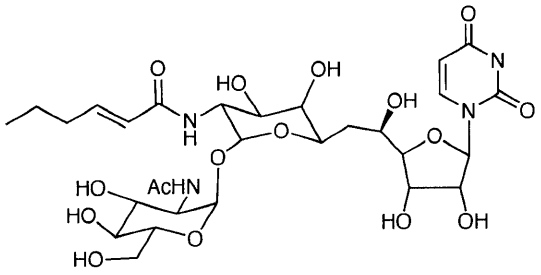


Figure 6-14: The structure of tunicamycin, a natural product found to inhibit eukaryotic *N*-linked glycosylation.

A panel of bisubstrate analogs was generated to test for PglB inhibition (Figure 6-15); these peptides were synthesized through modification of a key aspartate within the peptide scaffold in a similar approach to that taken for the extended asparagines illustrated in Figure 6-13. As shown in Figure 6-15, two distinct classes of molecules were designed as bisubstrate

analogs. In the first set (**40-44**), modified benzyl groups were coupled to the free aspartate to generate aryl asparagines. These analogs were envisioned to be potential inhibitors of PglB based on past analysis of the crystal structures of several carbohydrate-binding proteins, in which aromatic and hydrophobic amino acid residues line the binding groove.[35,36] Recent studies have shown that key interactions between these hydrophobic side chains and the pyranose rings of the glycan are often critical for substrate recognition.[37] We theorized that PglB may contain similar types of residues in the sugar-binding region, and the introduction of aryl rings into the active site would result in favorable hydrophobic interactions and/or π-stacking that might lead to inhibition.

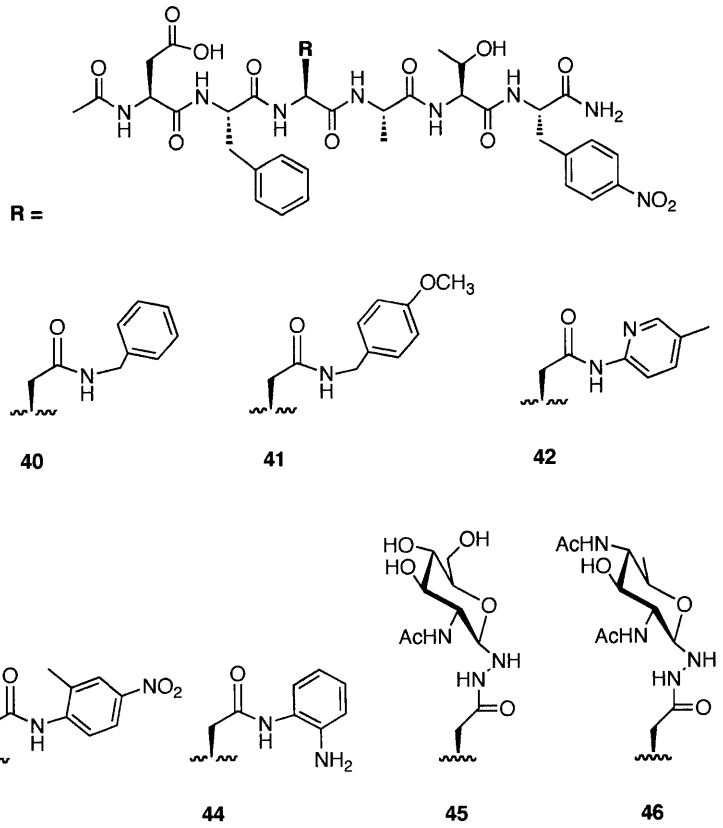


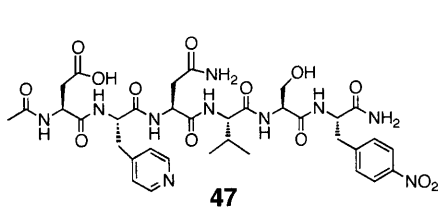
Figure 6-15: Aryl asparagines (**40-44**) and neoglycoconjugates (**45, 46**) outlined in this study.

In the second set of bisubstrate analogs, either *N*-acetylglucosamine (GlcNAc) or 2,4-diacetamidobacillosamine (Bac) was appended to the asparagine backbone to afford the neoglycoconjugates **45** and **46**. These molecules were synthesized using chemoselective ligation, in which the sugar aldehydes are coupled to the hydrazidoasparagine side chain (**14**) at low pH. As in the case of the aryl asparagines, it was postulated that the carbohydrate moieties appended to the inhibitor peptides would bind to regions of the enzyme active site that interact with the glycosyl donor substrates. However, no inhibition of PglB was observed for any of these peptides (**40-46**).

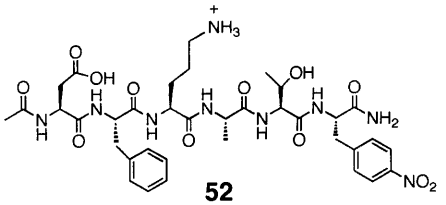
In addition to the asparagine isosteres and the bisubstrate analog peptides outlined in this chapter, many other peptides were synthesized and tested against PglB in an effort to observe inhibition; several of these peptides are depicted in Figure 6-16. These peptides include modifications at the Asn residue as well as the required Asp or Ser/Thr sites of the peptide scaffold. For example, phosphoserine (**54, 55**) was utilized to introduce an additional negative charge at these positions, as was a Cys residue to serve as a potential nucleophile (**49-51**). In one case, an *N*-methylserine residue was incorporated in the place of Ser/Thr (**56**) to prevent the peptide from adopting the Asx turn. This peptide was not glycosylated by PglB, thus highlighting the importance of this peptide conformation to glycosylation in *C. jejuni*.

In summary, a wide array of peptides has been synthesized and tested against PglB for inhibition. Despite the variety of peptides described herein, none exhibited inhibitory activity against PglB. These findings were surprising, particularly in the case of the asparagine isosteres, as several similar peptides have previously been found to be inhibitors of OT (Table 6-1). All peptides were additionally tested as possible substrates, and all except for one, hydroxyasparagine (**13**), were not glycosylated under the assay conditions employed. These

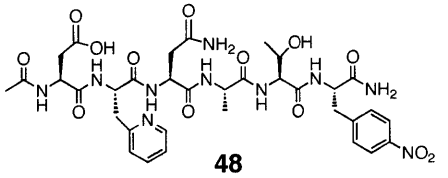
results indicate that PglB may display exquisite specificity for the glycan acceptor compared with OT; in support of this hypothesis, PglB exhibits a K_m = 1 μM for the acceptor peptide compared with that for OT (K_m = 240 μM for the tripeptide Bz-NLT).[21] While PglB is unable to tolerate even minor modifications to the asparagine amide, OT appears to bind a variety of asparagine isosteres and other peptide analogs. Further biochemical characterization of PglB is required to understand the source of this unique substrate specificity.



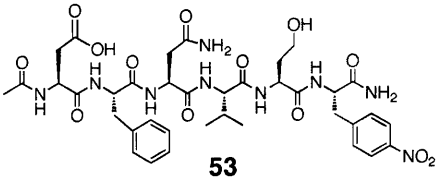
47



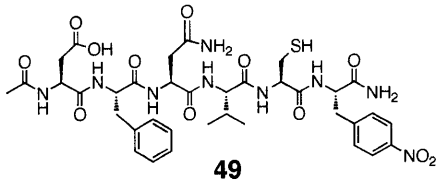
52



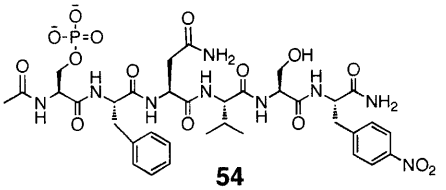
48



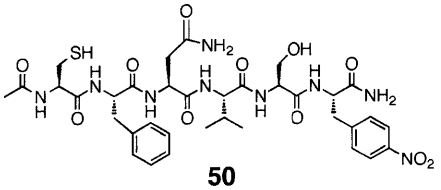
53



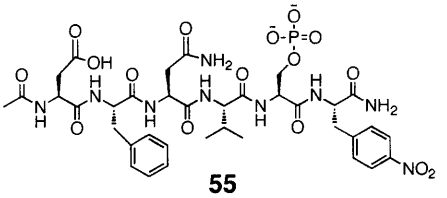
49



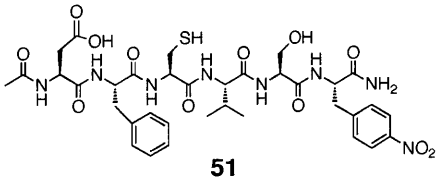
54



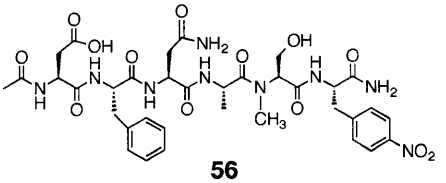
50



55

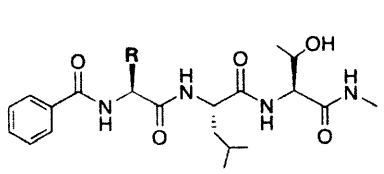


51

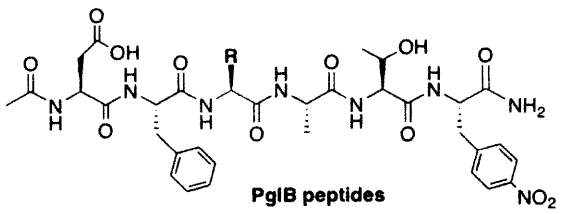


56

Figure 6-16: Representative examples of additional peptides designed as inhibitors of PglB.



OT peptides



PglB peptides

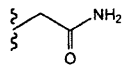
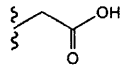
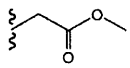
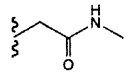
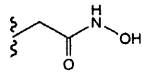
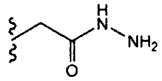
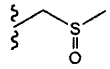
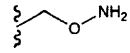
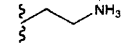
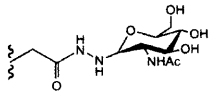
R	OT activity	PglB activity
	K_m = 0.24 mM	K_m = 1.0 μM
	No activity	No activity
	No activity	No activity
	K_i > 100 μM	No activity
	K_m = 20-50 μM	K_m > 100 μM
	K_i = 2.5 μM	No activity
	No activity	No activity
	No activity	No activity
	K_i = 40 μM	No activity
	K_i = 0.06 μM	No activity
	K_i = 4.5 μM	No activity

Table 6-1: Comparison of peptide-based inhibitors of OT (*S. cerevisiae*) and PglB (*C. jejuni*). In the case of PglB, all potential inhibitors were assayed at concentrations ranging from 50-250 μM.

Screening PglB with Small Molecule Alkylating Agents

Based on the lack of enzyme inhibition observed by the peptide-based inhibitors described in this chapter, PglB was screened against several small molecule alkylating agents. Although these agents do not mimic the peptide substrate, they can often provide information about key residues involved in substrate binding and/or catalysis by covalently modifying specific residues and potentially disrupting enzyme function. Earlier studies in the Imperiali group indicated that treatment of OT with the sulfhydryl modifying agent methyl methanethiolsulfonate (MMTS) resulted in complete inactivation of the enzyme.[38] Further analysis using a biotinylated analog of MMTS showed that alkylation occurred in the Wbp1p subunit, and that incubation of OT with the substrate dolichyl-pyrophosphate chitobiose prior to introduction of MMTS was sufficient to prevent alkylation, suggesting that the Wbp1p may be involved in binding the glycan donor.

PglB was incubated with three different alkylating agents, the cysteine-labeling agents MMTS, iodoacetamide (IAA), and *N*-ethylmaleimide (NEM), and the histidine-specific diethylpyrocarbonate (DEPC) at an initial concentration of 5 mM. After 30 mins, the acceptor peptide was introduced into the mixture to initiate glycosylation. Aliquots were removed from the reaction over time, partitioned between the aqueous and organic phases, and analyzed by scintillation counting. As indicated in Figure 6-17, the cysteine-labeling agents MMTS, IAA, and NEM had no effect on the rate of glycosylation. Surprisingly, total inactivation of PglB was achieved with DEPC.

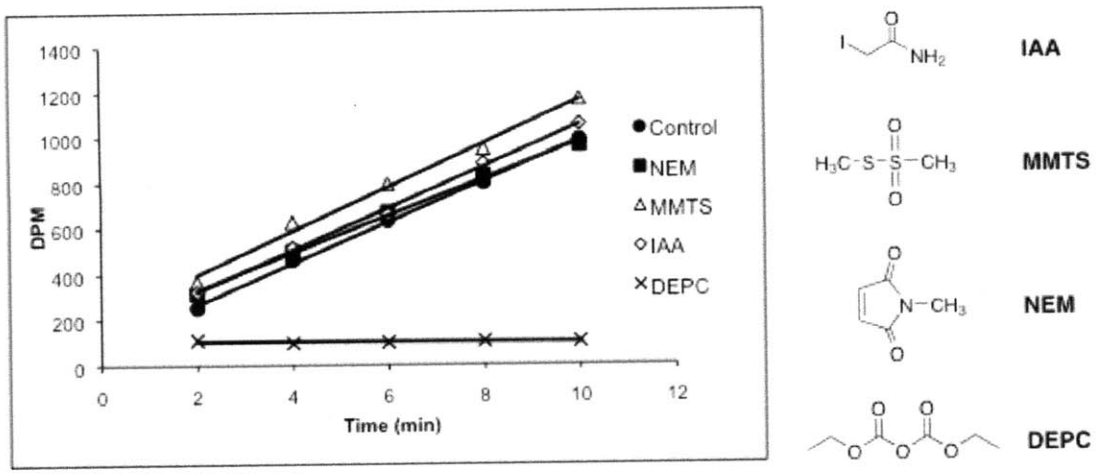


Figure 6-17: Activity of PglB in the presence of the Cys-alkylating agents NEM, MMTS, and IAA, and the His-alkylator DEPC. PglB was incubated with 5 mM of each alkylating agent for 30 mins prior to the addition of the acceptor peptide to initiate glycosylation.

Further investigation of this result indicated that enzyme inactivation was dependent on both time and DEPC concentration (Figure 6-18) and occurred regardless of whether PglB was introduced as pure protein or as part of a crude membrane fraction. DEPC has been previously shown to modify histidine over other amino acid residues in a specific manner.[39] In addition, it has been used to identify critical histidine residues in proteins where X-ray crystal structures were not readily available, as in the case of spinach leaf ferredoxin-$NADP^+$ oxidoreductase[40] and *E. coli* RNA polymerase.[41] In some cases, enzymatic activity of the DEPC-modified protein can be restored after treatment with hydroxylamine, which acts to remove the carboxyethyl group from the imidazole ring (Figure 6-19). However, efforts to restore the activity of PglB after DEPC incubation were unsuccessful due to the lability of Und-PP-Bac-[^{3}H]GalNAc; degradation of this substrate was rapidly observed at even low hydroxylamine concentrations.

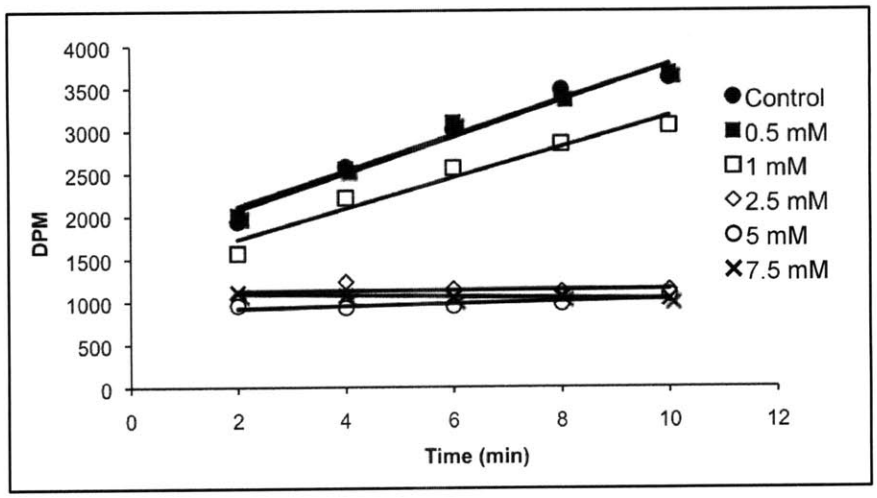


Figure 6-18: Concentration-dependent inactivation of PglB by the histidine alkylation agent, DEPC.

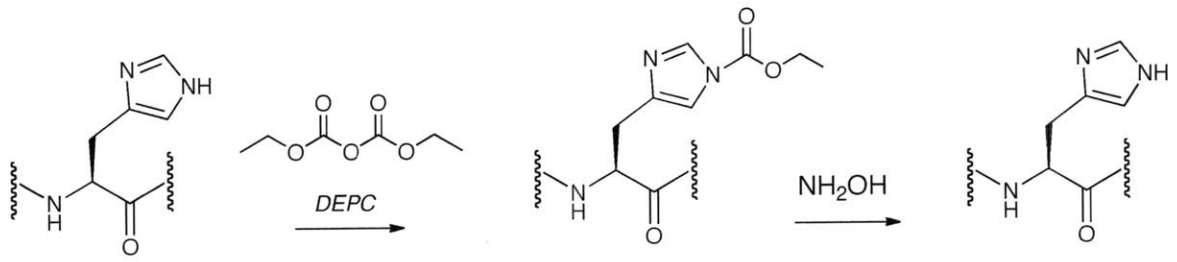


Figure 6-19: Alkylation of histidine residues by DEPC, which can be reversed through incubation with NH_2OH.

Functional Characterization of Histidine to Alanine Mutants of PglB

The inactivation of PglB by DEPC suggests that a histidine residue may play an important role in substrate binding or catalysis. Interestingly, this finding is supported by the observation that a single point mutation of PglB in which His479 is replaced with Ala results in a complete loss of enzymatic activity relative to the wild type enzyme.[42] In order to further characterize the role of His479 in the activity of PglB, the remaining histidine residues in the

PglB sequence were systematically mutated to alanine and these mutant enzymes were tested for activity. As shown in Figure 6-20, PglB contains a total of five histidine residues, four of which are found in the predicted N-terminal transmembrane domain of the enzyme. The fifth histidine residue, His479, is located at the beginning of the large soluble C-terminal domain in close proximity to the canonical WWDXGX motif.

MLKKEYLKNPYLVLFAMIVLAYVFSVFCRFYWVWWASEFNEYFFNNQLMIISNDGYAFAEGARDMIAGF
HQPNDLSYYGSSLSTLTYWLYKITPFSFESIILYMSTFLSSLVVIPIILLANEYKRPLMGFVAALLASV
ANSYYNRTMSGYYDTDMLVIVLPMFILFFMVRMILKKDFFSLIALPLFIGIYLWWYPSSYTLNVALIGL
FLIYTLIFHRKEKIFYIAVILSSLTLSNIAWFYQSAIIVILFALFALEQKRLNFMIIGILGSATLIFLI
LSGGVDPILYQLKFYIFRSDESANLTQGFMYFNVNQTIQEVENVDFSEFMRRISGSEIVFLFSLFGFVW
LLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMALGFGFLLSEFKAILVKKYSQLTSNVCIVFATI
LTLAPVFIHIYNYKAPTVFSQNEASLLNQLKNIANREDYVVTWWDYGYPVRYYSDVKTLVDGGKHLGKD
NFFPSFSLSKDEQAAANMARLSVEYTEKSFYAPQNPILKSDILQAMMKDYNQSNVDLFLASLSKPDFKI
DTPKTRDIYLYMPARMSLIFSTVASFSFINLDTGVLDKPFTFSTAYPLDVKNGEIYLSNGVVLSDDFRS
FKIGDNVVSVNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFNSAYVQMFFLG
NYDKNLFDLVINSRDAKVFKLKI

Figure 6-20: Protein sequence of PglB. The N-terminal transmembrane domain (as predicted by TMHMM)[43] is highlighted in bold and underlined, and the WWDXGX motif conserved in all Stt3 homologs is shown in cyan. Histidine residues are colored in pink, and the DXXK motif identified by Kohda et al is in green.[44]

Plasmids encoding the desired PglB mutants were prepared using standard site-directed mutagenesis techniques, and these constructs were transformed into *E. coli* BL21 (DE3) RIL cells and overexpressed. Each mutant was then prepared as a crude cell membrane fraction and analyzed by Western blot using an anti-T7 antibody directed against the N-terminal T7 tag, which showed that all mutants were expressed at roughly the same level. These mutants were then tested for glycosylation activity using the standard assay, as shown in Figure 6-21. Interestingly, all of the PglB mutants exhibited normal level of activity except for His479Ala, suggesting that perhaps His479 plays an important role in substrate binding or catalysis.

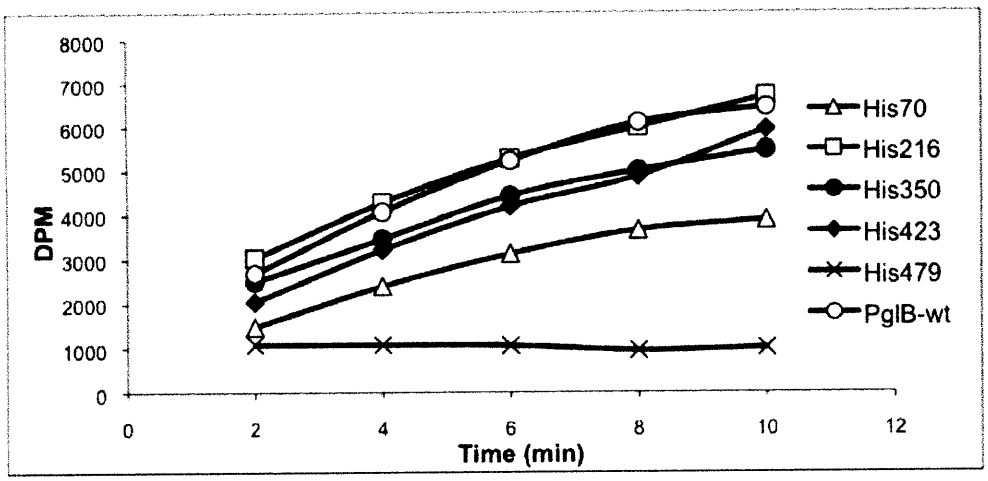


Figure 6-21: Activity of PglB alanine mutants.

In December 2009, a preliminary crystal structure of the soluble C-terminal domain of PglB was published by Kohda et al (Figure 6-22).[45] It is unclear whether this structure represents a biologically relevant form of PglB, as the C-terminal domain is unable to catalyze substrate turnover. In addition, the structure is missing several large sections of the soluble domain, and the lysine residues of the protein were methylated following purification to aid crystallization. Nonetheless, analysis of this structure reveals that His479 is situated in close proximity to the conserved WWDXGX motif. Thus, it follows that DEPC modification of this particular histidine residue could affect substrate binding or catalysis either directly, by preventing this residue from participating in the reaction mechanism, or indirectly, through structural perturbation of the critical WWDXGX motif thus altering its proper function. Regardless, further work is required to determine whether this key histidine residue is involved in the catalytic mechanism of PglB.

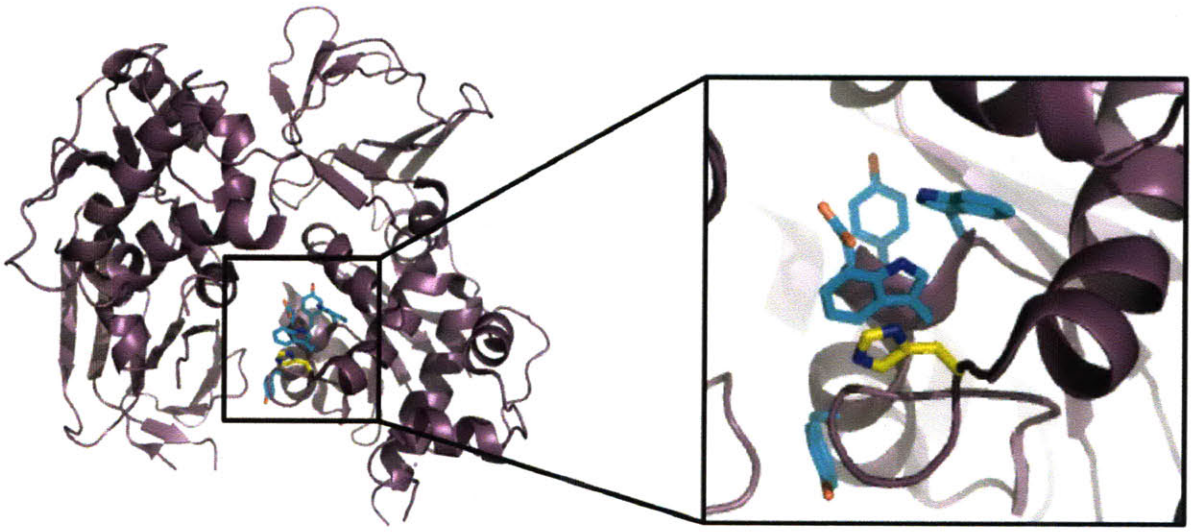


Figure 6-22: Crystal structure of the C-terminal soluble domain of PglB (PDB code: 3AAG).[45] His479 is highlighted in yellow, illustrating its proximity to the WWDYGY motif (cyan).

In addition to the histidine mutants described above, initial work was carried out to investigate the importance of a newly identified motif in oligosaccharyl transferase homologs, DXXK. In a recent report, Kohda et al assert that this sequence is highly conserved among all known OTases, and present genetic knockout experiments in the hyperthermophile *Pyrococcus furiosus* to highlight its importance in glycosylation.[44] To this end, three PglB mutants were prepared to further analyze the role of these residues, the two single mutants (D519A and K522A), as well as the double mutant (D519A/K522A). The mutants were generated in the same manner as previously outlined, and analysis of crude cell membrane fractions indicated that the overall expression level of each protein was roughly comparable. However, functional analysis of these mutants indicated that they were all catalytically competent (Figure 6-23). Although these results were surprising, examination of the amino acid sequence of the PglB C-terminal domain reveals that in fact three different DXXK motifs are present; thus, perhaps one of the remaining two motifs could in fact be key sequence. Current work in the laboratory is

underway to investigate these further investigate this motif in PglB to determine if in fact it is important for oligosaccharyl transferase activity in vitro.

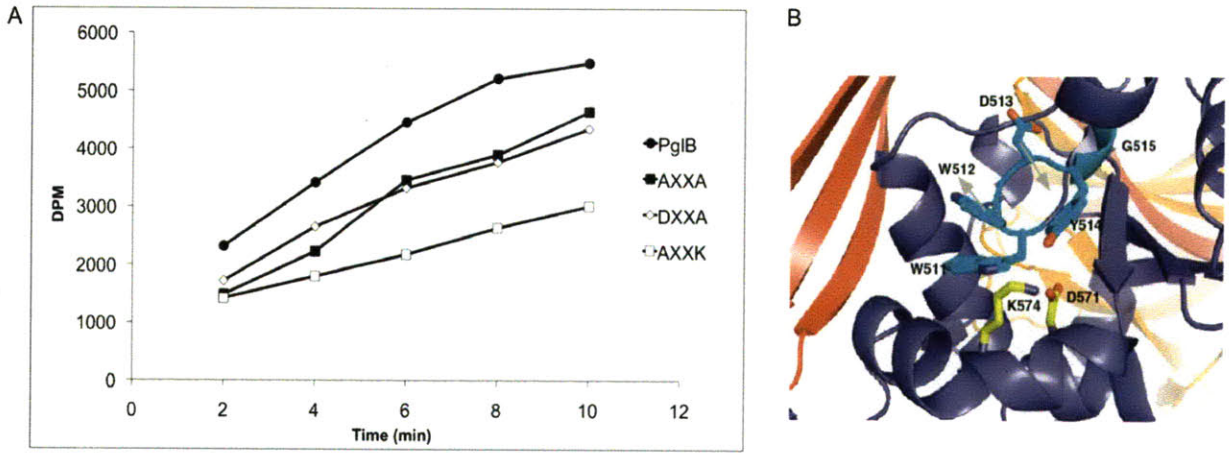


Figure 6-23: (A) PglB activity assay, comparing the overall reaction rates of the DXXK mutants; (B) Crystal structure of the C-terminal soluble domain of the *P. furiosus* OTase, highlighting the proximity of the newly identified DXXK motif (yellow) to the WWDYG sequence (cyan). Figure adapted from Kohda et al.[44]

Conclusions

In this chapter, the design, synthesis, and in vitro screening of a large panel of substrate-based peptides was described in an effort to obtain an inhibitor of the bacterial oligosaccharyl transferase PglB. However, despite trying a wide variety of reaction conditions, no inhibition of the enzyme was ever observed. It is currently unclear why we were unable to see inhibition of PglB, particularly in light of the fact that several similar peptides were capable of affecting activity of the yeast OT. One explanation for this difference may be related to the affinity of each enzyme for its substrate; in the case of OT, the K_m = 240 μM for its substrate peptide, while PglB exhibited a K_m = 1 μM which represents a remarkable 200-fold variation. This

marked difference in K_m implies that PglB binds its substrate with greater affinity, and may suggest a concomitant increase in peptide specificity not observed for OT. Another hypothesis involves a recent finding by the Szymanski laboratory, in which PglB was found to be involved in the hydrolysis of the Und-PP-BacGalNAc$_5$Glc glycan donor to provide free oligosaccharide.[46] The levels of free oligosaccharide in the periplasm of *C. jejuni* have been linked to changes in growth conditions and osmolarity of the surrounding environment, and are proposed to play a protective role for the organism. Although further work is required to fully understand how this free oligosaccharide is formed and whether PglB is directly involved, perhaps our inability to observe inhibition of PglB is somehow related to in vitro formation of the hydrolyzed glycan.

In addition, this chapter describes efforts to identify key amino acid residues within PglB to gain insight into substrate binding and/or catalysis. To this end, PglB was incubated with a small panel of enzyme alkylating agents and found to be completely inactivated by the histidine-specific DEPC in a time and concentration dependent manner. In order to further investigate this finding, a series of histidine mutants of PglB were generated; it was found that only mutation of His479, located near the putative enzyme active site, had an effect on enzyme activity. Enzyme mutants were also generated to explore a newly identified DXXK motif. Future work in this area will entail further investigation of PglB using site-directed mutagenesis, and may provide key insight into the elusive reaction mechanism of oligosaccharyl transferase.

Acknowledgements

I am grateful to Professor Eranthie Weerapana for early help with peptide design and synthesis and the oligosaccharyl transferase assay. Special thanks to Professor K. Jebrell Glover for his advice on handling and purification of PglB, and to Drs. Matthieu Sainlos and Langdon Martin for their assistance with MALDI TOF MS analysis.

Experimental Methods

General information

Unless otherwise noted, all solvents and reagents were obtained commercially and used without further purification. Fmoc-protected amino acids and coupling agents were purchased from Novabiochem (Gibbstown, NJ) and GenScript (Piscataway, NJ). Solvents were obtained from VWR (West Chester, PA) or Sigma Aldrich (St. Louis, MO). Fmoc-PAL-PEG-PS resin was acquired from Applied Biosystems (Carlsbad, CA). Oligonucleotides were purchased from Eurofin MWG Operon (Huntsville, AL) or Sigma Life Sciences (St. Louis, MO). Sequencing of all bacterial plasmids was conducted by the MIT CCR Biopolymers Laboratory (Cambridge, MA). Analytical TLC was performed on silica gel 60 F254 plates (EMD Chemicals) and visualized with either UV or staining with ninhydrin, vanillin, or CAM. Flash column chromatography was performed using forced flow of the indicated solvent on silica gel 60, particle size 0.043-0.063 mm (EMD Chemicals). HPLC was carried out on a Waters Prep LC 4000 or Waters Delta 600 systems, each equipped with a Waters 2487 dual wavelength detector. For analytical HPLC, a C_{18} YMC ODS-A 5 μm (4.6 x 250 mm) column was used, and for preparatory HPLC, a C_{18} YMC-Pack ODS-A 5 μm (250 x 20 mm) column was employed. ESI-

MS was conducted on a Mariner instrument (Applied Biosystems), and MALDI-TOF MS was performed on the Voyager system (Applied Biosystems). NMR spectra were acquired on a Bruker 400 MHz spectrometer equipped with a 5 mm variable temperature gXH probe. Chemical shifts (δ) are reported in parts per million (ppm) and referenced to internal standards, where either the HOD signal at 4.80 ppm, $C\underline{H}Cl_3$ signal at 7.28 ppm, or $(C\underline{H}_3)_2SO$ was used for ^1H analysis. Coupling constants (*J*) are reported in Hertz (Hz) and multiplicities are abbreviated as singlet (s), broad singlet (br s), doublet (d), triplet (t), multiplet (m), doublet of doublets (dd), and doublet of triplets (dt). Infrared spectrophotometry was carried out on a Perkin-Elmer Model 2000 instrument with samples dried on NaCl plates (25 mm diameter x 4 mm width, Sigma Aldrich).

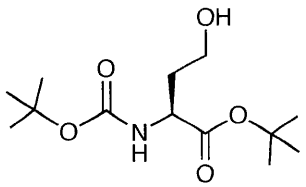
Peptide Synthesis

All peptides were synthesized using standard Fmoc-based solid phase protocols on Fmoc-PAL-PEG-PS resin (0.19 mmol/g). The resin was swelled first in CH_2Cl_2 (5 mins) then DMF (5 mins) prior to synthesis. Deprotection of the terminal Fmoc group was achieved by exposure of the resin to an excess of 20% 4-methylpiperidine in DMF (3 x 5 mins). Each amino acid was coupled to the resin by incubation with the amino acid (4 equiv), PyBOP (4 equiv), and DIPEA (8 equiv) for 45 mins. Between each deprotection and coupling step, the resin was washed with DMF (5 x 1 min) then CH_2Cl_2 (5 x 1 min). The TNBS test was used to check the resin for coupling efficiency, in which the formation of red color indicates the presence of a free amine.[47] All peptides were acetylated at the N-terminus by incubation with acetic anhydride (10 equiv) and pyridine (10 equiv) in DMF for 30 mins. The side chain protecting groups were removed and the peptides cleaved from the resin by exposure to a TFA:CH_2Cl_2:H_2O:TIS mixture

(90:5:2.5:2.5) for 3 hrs while shaking vigorously, followed by filtration to remove the beads and solvent evaporation under a stream of nitrogen. The resulting pellet was triturated with cold Et_2O, and the peptides were purified by preparative C_{18} RP-HPLC using a 45 minute gradient of 5-95% mobile phase B, where the mobile phases are H_2O / 0.1% TFA (A) and CH_3CN / 0.1% TFA (B). Purified peptides were characterized by analytical RP-HPLC for purity and either ESI-MS or MALDI-TOF MS for identity. Peptides were stored as solutions of DMSO and quantified by UV absorbance using the ε_{280} of Nph (12,500 $M^{-1}cm^{-1}$).

Synthesis of L-Oxonorvaline

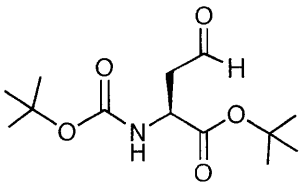
The methyl ketone amino acid was synthesized following previously reported procedures with slight modifications.[24,25]



25

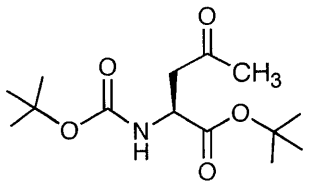
a. (L)-Boc-Homoserine-OtBu (**25**): To a stirring solution of Boc-Asp-OtBu (**24**, 800 mg, 2.76 mmol) in 10 mL dry THF under an atmosphere of argon at -15 °C, triethylamine (385.2 μL, 2.76 mmol) and ethyl chloroformate (263.3 μL, 2.76 mmol) were added. After 30 mins, the tetraammonium chloride was filtered from solution, and the filtrate was added dropwise to a stirring solution of $NaBH_4$ (209.12 mg, 5.52 mmol) in H_2O (10 mL) at 10 °C over the course of 10 mins. After stirring at room temperature for 4 hrs, the solution was acidified to pH 6.5 with 1 M HCl. The mixture was extracted with EtOAc (2 x 20 mL), and the combined EtOAc layers

were then washed with 10% NaOH, H_2O, and brine (20 mL each). The EtOAc solution was dried (Na_2SO_4) and then concentrated under vacuum to yield the desired product as a light yellow oil with no discernible impurities (~ 750 mg, 92%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ 1.50 (18H, s, C*H*$_3$), 2.2 (2H, m, C*H*$_2$CH$_2$OH), 3.72 (2H, m, CH$_2$C*H*$_2$OH), 4.43 (1H, m, α-C*H*), 5.41 (1H, d, N*H*). ESI-MS: calcd for $[C_{13}H_{25}NO_5 + H]^+$ requires *m/z* = 275.2, found *m/z* = 275.3.



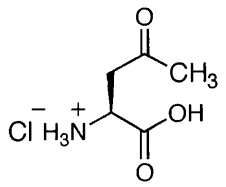
26

b. (L)-Boc-β-Aspartyl-semialdehyde-OtBu (**26**): Anhydrous DMSO (541.8 μL, 7.62 mmol) was added to a stirring solution of oxalyl chloride (1.91 mL, 3.81 mmol) in dry CH_2Cl_2 at -78 °C under an atmosphere of argon. To this mixture, Boc-homoserine-OtBu (**25**, 350 mg, 1.27 mmol) dissolved in dry CH_2Cl_2 (4.5 mL) was added and allowed to stir for 30 mins, followed by the addition of anhydrous DIPEA (2.21 mL, 12.7 mmol). After 5 mins, the reaction was removed from the cold bath and allowed to warm to room temperature while stirring over the course of 2 hrs. The solution was then diluted with CH_2Cl_2 (30 mL) and extracted with brine (2 x 30 mL). The organic layer was dried over Na_2SO_4, concentrated under vacuum, and purified by silica gel column chromatography (1:1 hexanes:EtOAc) to provide the pure product, **26** (295 mg, 85%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ 1.49 (18H, s, C*H*$_3$), 2.93 (2H, m, CHC*H*$_2$C(O)H), 4.48 (1H, m, α-C*H*), 5.40 (1H, d, N*H*) 9.71 (1H, s, C(O)*H*). ESI-MS: calcd for $[C_{13}H_{23}NO_5 + H]^+$ requires *m/z* = 273.2, found *m/z* = 273.2.



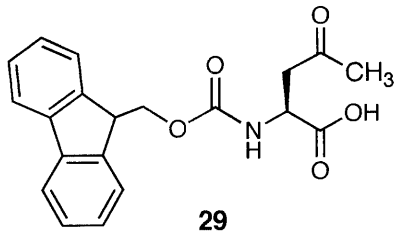
27

c. (L)-Boc-Oxonorvaline-OtBu (**27**): Diazomethane was generated from Diazald (3.26 g, 15.22 mmol) in a mixture of ether (12 mL), KOH (3.6 g), EtOH (4 mL), and H_2O (3 mL) at 65 °C, using a ClearSeal distillation setup and following standard procedures. Caution was taken during the distillation due to the extreme volatility of diazomethane. The freshly distilled diazomethane was slowly dripped into a stirring solution of Boc-β-aspartyl-semialdehyde-OtBu (**26**, 300 mg, 1.10 mmol) in dry CH_2Cl_2 (2 mL) at -78 °C as it was formed. The solution was allowed to stir overnight, after which the characteristic yellow color of diazomethane disappeared. The reaction was then quenched with 0.1 M HCl (10 mL) and extracted with CH_2Cl_2 (3 x 10 mL). The organic layers were combined, dried over Na_2SO_4, and purified by silica gel chromatography (1.5:1 hexanes:EtOAc) to afford the desired product (287 mg, 91%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ 1.46 (18H, s, C*H*$_3$), 2.11 (3H, C(O)C*H*$_3$), 2.89 (1H, dd, J = 4.1, J = 12.8, CHC*H*$_2$C(O)CH$_3$), 3.10 (1H, dd, J = 4.1, J = 12.8, CHC*H*$_2$C(O)CH$_3$), 4.35 (1H, m, α-C*H*), 5.42 (1H, d, J = 6.7, N*H*). ESI-MS: calcd for $[C_{14}H_{25}NO_5 + H]^+$ requires m/z = 287.2, found m/z = 287.3.



28

d. (L)-Oxonorvaline-OH hydrochloride (**28**): Boc-oxonorvaline-OtBu (**27**, 250 mg, 0.87 mmol) was added to a stirring solution of CH_2Cl_2:TFA (2:1, 6 mL) and incubated for 16 hrs at room temperature to yield a deep red mixture. The reaction was then concentrated under vacuum, and the remaining red oil was washed with Et_2O (2 x 10 mL) and lyophilized from H_2O to afford an orange solid as a TFA salt. The salt was dissolved in MeOH (6 mL) and incubated with Dowex resin (Cl^- form, 0.5 g). The suspension was stirred at room temperature for 24 hrs, and the resin was filtered and washed with MeOH (3 x 10 mL). The combined organic fractions were concentrated under vacuum and the resulting orange solid was recrystallized from MeOH:EtOAc (1:1) to achieve the desired product **28** (110 mg, 76%). ^{1}H-NMR (400 MHz, D_2O): δ 2.06 (s, 3H, $CH_2C(O)CH_3$), 3.18 (2H, dd, J = 3.9, J =12.8, $CH_2C(O)CH_3$), 4.16 (1H, t, J = 3.9, α-C*H*). ESI-MS: calcd for $[C_5H_{10}NO_3 + H]^+$ requires m/z = 132.1, found m/z = 132.1.



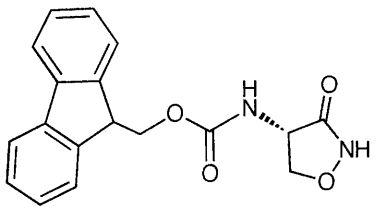
29

e. (L)-Fmoc-Oxonorvaline-OH (**29**): A solution of oxonorvaline-OH (**28**, 100 mg, 0.978 mmol) and $NaHCO_3$ (458.7 mg, 5.46 mmol) in H_2O was prepared in a round bottom flask. 9-Fluorenylmethyl *N*-succinimidyl carbonate (Fmoc-OSu, 398.0 mg, 1.18 mmol) dissolved in

dioxane (8 mL) was then added to the stirring solution in a dropwise fashion. The pH of the cloudy suspension was adjusted to 8.5 by the addition of additional $NaHCO_3$ (~50 mg) and after 3 hrs, TLC analysis (1:4 MeOH:CH_2Cl_2, ninhydrin stain) indicated complete conversion of the starting material. The reaction mixture was then extracted with Et_2O (3 x 10 mL) and the aqueous layer acidified to pH 2.5 by the slow addition of 1 M HCl, causing a thick white precipitate to form. The precipitate was dissolved in EtOAc (10 mL), and the aqueous layer was extracted with additional EtOAc (2 x 10 mL). The EtOAc layers were combined, dried over Na_2SO_4, and purified by flash silica chromatography (5% MeOH/CH_2Cl_2) to yield the desired Fmoc-protected amino acid **29** (74%). ^{1}H-NMR (400 MHz, $(CD_3)_2SO$): δ 2.07 (3H, s, $CH_2C(O)CH_3$), 4.32 (1H, m, C*H*-Fmoc), 4.40 (1H, m, CH_2-Fmoc), 4.46 (1H, m, CH_2-Fmoc), 4.63 (1H, m, α-C*H*), 7.39 (2H, m, Ar), 7.43 (2H, m, Ar), 7.61 (1H, d, J = 6.1, N*H*), 7.69 (2H, d, J = 7.0, Ar), 7.86 (2H, m, Ar) (Appendix 1). ESI-MS: calcd for $[C_{20}H_{19}NO_5 + H]^+$ requires m/z = 353.1, found m/z = 353.3.

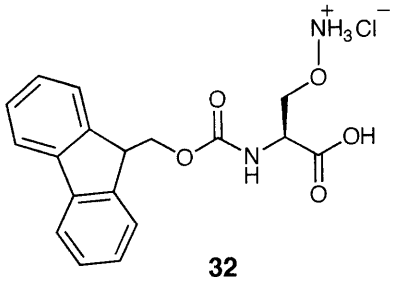
Synthesis of Aminoserine

The aminoserine amino acid was prepared using previously published protocols with slight modification.[27]



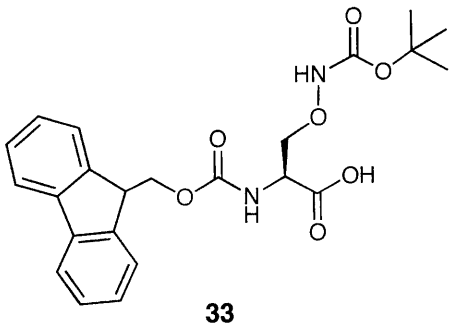
31

a. (L)-Fmoc-Cycloserine (**31**): L-Cycloserine (100 mg, 0.978 mmol) was protected with Fmoc in a similar procedure to that described above for Fmoc-oxonorvaline-OH (**29**) to yield the desired Fmoc-cycloserine (**31**) in good yield. ^{1}H-NMR (400 MHz, $CDCl_3$): δ 4.12 (1H, t_{app}, J = 7.1, CH_2ON), 4.25 (1H, t, CH-Fmoc), 4.45 (2H, d, J = 4.2, CH_2-Fmoc), 4.69 (1H, m, CH_2ON), 4.83 (1H, t_{app}, J = 7.8, α-CH), 5.45 (1H, br s, NH), 7.30 (2H, t, J = 7.1, Ar), 7.42 (2H, t, J = 7.1, Ar), 7.60 (2H, d, J = 7.0, Ar), 7.81 (2H, d, J = 7.7, Ar), 8.05 (1H, m, NH). ESI-MS: calcd for $[C_{18}H_{16}N_2O_4 + H]^+$ requires m/z = 324.3, found m/z = 324.4.



32

b. (L)-Fmoc-Aminoserine Hydrochloride (**32**): To a stirring solution of **31** (136 mg, 0.42 mmol) in THF (2 mL) at 60 °C, 2 M HCl (2 mL) was added in a slow, dropwise manner. The reaction was monitored by TLC, and after 96 hrs was neutralized by the addition of $NaHCO_3$ and concentrated under vacuum to provide the crude product (**32**), which was purified by preparative HPLC in high yield (87%). ^{1}H-NMR (400 MHz, $(CD_3)_2SO$): δ 4.22 (6H, m, CH-Fmoc, CH_2-Fmoc, α-CH, CH_2ON), 7.27 (2H, t, J = 7.1, Ar), 7.36 (2H, t, J = 7.1, Ar), 7.63 (2H, d, J = 7.0, Ar), 7.84 (2H, d, J = 7.6, Ar). ESI-MS: calcd for $[C_{18}H_{18}N_2O_5 + H]^+$ requires m/z = 342.4, found m/z = 324.5.

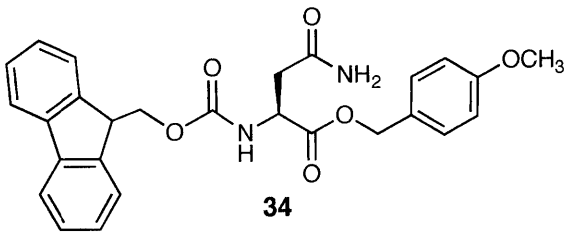


33

c. (L)-Fmoc-Aminoserine(Boc)-OH (**33**): Fmoc-Aminoserine HCl (**32**, 100 mg, 0.31 mmol) dissolved in DMF (5 mL) was treated with DIPEA (108.0 μL, 0.62 mmol) and di-*tert*-butyl dicarbonate (Boc_2O, 74.2 mg, 0.34 mmol) while stirring. The reaction was incubated for 3 hrs, and the pH was monitored over the course of the reaction, adding DIPEA to ensure a pH of 8.5. Once the reaction was complete, the reaction was diluted with Et_2O (10 mL) and washed with 0.1 M HCl (2 x 10 mL), H_2O (2 x 10 mL), and 2% $NaHCO_3$ (3 x 10 mL). The combined aqueous layers were then acidified to pH 2 with 1 M HCl, and the precipitated material was dissolved in ether, dried ($NaSO_4$), and concentrated under vacuum. The crude product was recrystallized from acetone/hexanes to give the final product, **33**. ^{1}H-NMR (400 MHz, $(CD_3)_2SO$): δ 1.42 (9H, s, C*H*$_3$), 3.91 (1H, dd, *J* = 6.5, C*H*$_2$ON), 4.01 (1H, m, C*H*$_2$ON), 4.21 (4H, m, C*H*-Fmoc, C*H*$_2$-Fmoc, and α-C*H*), 7.23 (2H, t, *J* = 7.3, Ar), 7.39 (2H, t, *J* = 7.3, Ar), 7.57 (1H, d, *J* = 6.0, N*H*), 7.67 (2H, d, *J* = 7.2, Ar), 7.86 (2H, d, *J* = 7.6, Ar), 10.03 (1H, br s, COO*H*). ESI-MS: calcd for calcd for $[C_{23}H_{26}N_2O_7 + H]^+$ requires *m/z* = 442.5, found *m/z* = 442.5.

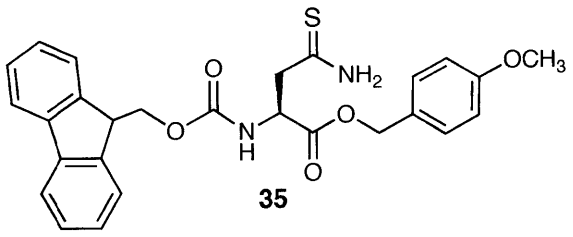
Synthesis of (L)-Fmoc-thioasparagine(xanthen-9-yl)-OH

The synthesis of the desired thioasparagine amino acid, Fmoc-thioasparagine-(xanthen-9-yl)-OH, (**37**), was achieved using previously published procedures with slight modification.[48]



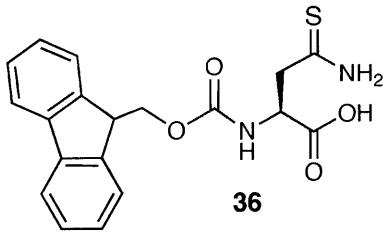
34

a. (L)-Fmoc-asparagine 4-methoxybenzyl ester (**34**): A solution of (L)-Fmoc-Asn-OH (1.5 g, 4.2 mmol, **18**), 4-methoxybenzyl chloride (688.9 uL, 5.08 mmol), sodium iodide (919.4 mg, 6.13 mmol), and DIPEA (1.47 mL, 8.46 mmol) in dimethylacetamide (DMA, 30 mL) was prepared in a round bottom flask and stirred under nitrogen at room temperature for 18 hrs. After dilution with $CHCl_3$ (200 mL) and extraction with H_2O (2 x 200 mL), the organic layer was dried ($MgSO_4$) and filtered before removal of the solvent by evaporation. The solid was recrystallized from $CHCl_3$-pentane to give the desired product as a pale yellow solid (1.49 g, 71%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 2.70 (1H, dd, J = 4.6, J = 12.3, CHCH_2C(O)NH_2), 3.05 (1H, dd, J = 4.6, J = 12.3, CHCH_2C(O)NH_2), 3.75 (3H, s, CH_3OPh), 4.22 (1H, m), 4.35 (1H, m), 4.44 (1H, m), 4.63 (1H, m, α-CH), 5.18 (2H, s, CH_3OPhCH_2), 5.46 (1H, br s, C(O)NH_2), 5.63 (1H, br s, C(O)NH_2), 6.10 (1H, d, J = 6.4, Fmoc-NH), 6.87 (2H, d, J = 7.2, Ar), 7.32 (4H, m, Ar), 7.43 (2H, t, J = 7.5, Ar), 7.62 (2H, d, J = 7.0, Ar), 7.82 (2H, d, J = 7.3, Ar). ESI-MS: calcd for $[C_{27}H_{26}N_2O_6 + H]^+$ requires m/z = 475.2, found m/z = 475.2.



35

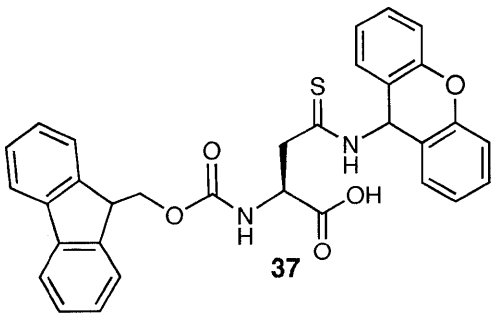
b. (L)-Fmoc-thioasparagine 4-methoxybenzyl ester (**35**): A mixture of Fmoc-Asn-OPMB (**34**, 1.2 g, 2.5 mmol) and phosphorus pentasulfide (632 mg, 2.78 mmol) in dry THF (25 mL) was assembled in a round bottom flask under an atmosphere of nitrogen and placed in a sonication bath for 3 hours, maintaining the bath < 40 °C. The mixture was filtered to remove precipitate, which was washed with hot $CHCl_3$ (3 x 15 mL). The $CHCl_3$ washes were combined with the THF filtrate and dried under vacuum. The crude product was purified by flash column chromatography (CH_2Cl_2:pentane:EtOAc; 3:3:2) to give the desired product (**35**) as a white solid (749 mg, 61%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 3.27 (2H, m, $CHCH_2C(S)NH_2$), 3.80 (3H, s, CH_3OPh), 4.22 (1H, m, C*H*-Fmoc), 4.43 (2H, d, J = 6.5, CH_2-Fmoc), 4.71 (1H, m, α-C*H*), 5.15 (2H, s, CH_2OPh-CH_2), 6.10 (1H, d, J = 5.8, N*H*-Fmoc), 6.87 (2H, d, J = 8.5, Ar), 7.32 (4H, m, Ar), 7.42 (2H, t, J = 7.5), 7.61 (2H, J = 7.1, Ar), 7.76 (2H, d, J = 7.6, Ar). ESI-MS: calcd for $[C_{27}H_{26}N_2O_5S + H]$+ requires m/z = 491.2, found m/z = 491.2.



36

c. (L)-Fmoc-thioasparagine-OH (**36**): Fmoc-ThioAsn-OPMB (**35**, 100 mg, 0.2 mmol) was treated with a solution of TFA (1 mL) containing 1% of both thioanisole and ethanedithiol

(EDT) for 30 mins at room temperature. The reaction was stopped upon addition of a Et_2O-pentane mixture (1:1, 10 mL), wherein a white precipitate was formed. The mixture was allowed to stand for 2 hr at 0 °C, after which it was transferred to a 15 mL Falcon tube and centrifuged at 3000 rpm for 5 mins. After removal of the supernatant by decantation, the precipitate was washed with Et_2O (2 x 5 mL) to remove impurities, affording the desired compound **36** as a white solid (47 mg, 63%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 2.85 (1H, dd, J = 4.6, J = 12.9, CHCH_2C(S)NH_2), 2.98 (1H, dd, J = 4.6, J = 12.9, CHCH_2C(S)NH_2), 4.36 (3H, m, C*H*- and CH_2-Fmoc), 4.67 (1H, m, α-C*H*), 7.33 (2H, t, J = 6.8, Ar), 7.42 (2H, t, J = 7.1, Ar), 7.64 (1H, d, J = 7.9, Ar), 7.70 (2H, d, J = 7.0, Ar), 7.90 (2H, d, J = 7.3, Ar), 9.22 (1H, br s, NH_2), 9.56 (1H, br s, NH_2). ESI-MS: calcd for $[C_{19}H_{18}N_2O_4S + H]^+$ requires m/z = 371.1, found m/z = 371.1.



37

d. (L)-Fmoc-thioasparagine-(xanthen-9-yl)-OH (**37**): A solution of Fmoc-thioasparagine-OH (**36**, 70 mg, 0.19 mmol) and 9-hydroxyxanthene (75.1 mg, 0.38 mmol) in glacial acetic acid (2.5 mL) was prepared in a round bottom flask and heated to 85 °C under an atmosphere of nitrogen for 2 hrs. Following the removal of the solvent by rotary evaporation, the orange-colored residue was triturated with Et_2O (3 x 3 mL) to provide the product as a white precipitate. The

product was recrystallized from $CHCl_3$ to give the final compound (48 mg, 46%). ^1H NMR (400 MHz, $(CD_3)_2SO$): δ 2.75 (1H, m, $CHCH_2C(S)NH_2$), 3.31 (1H, m, $CHCH_2C(S)NH_2$), 3.71 (1H, m, C*H*-Fmoc), 4.27 (2H, m, CH_2-Fmoc), 4.45 (1H, m, C*H*-Xan), 4.69 (H, m, α-C*H*), 7.08 (2H, m, Ar), 7.27 (2H, m, Ar), 7.33 (2H, t, $J = 6.1$, Ar), 7.43 (2H, t, $J = 6.9$ Hz), 7.50 (2H, t, $J = 7.7$, Ar), 7.69 (2H, m, Ar), 7.72 (1H, m, N*H*), 7.90 (2H, m, Ar), 8.02 (1H, d, $J = 7.1$, N*H*), 8.22 (2H, dd, $J = 3.1$, $J = 7.6$, Ar), 11.28 (1H, br s, CO_2H). FT-IR: ν_{max} (cm^{-1}) = 3310, 1730, 1650, 1515, 1250, 1170, 1030, 735. ESI-MS: calcd for $[C_{32}H_{26}N_2O_5S + H]^+$ requires *m/z* = 551.2, found *m/z* = 551.2.

Synthesis of Thiosasparagine peptide, Ac-Asp-Phe-ThioAsn-Val-Ser-Nph-NH_2 (**17**)

The Ac-Asp-Phe-Asn-Val-Ser-Nph-NH_2 peptide was synthesized on PAL-PEG resin following standard solid-phase protocols. After deprotection and cleavage from the resin, the peptide was purified by RP-HPLC, lyophilized, and characterized by MALDI TOF MS. The peptide (0.38 μmol) was then dissolved in DMSO and added to a solution of Lawesson's reagent (0.5 mg, 1.24 μmol) in THF:DMSO (1:1, 250 μL). The mixture was exposed to sonication at 25 °C in an H_2O bath for 6 hrs, after which the reaction was purified by HPLC to provide the desired product (**17**) in a very poor yield (~ 8%). ESI-MS: calcd for $[C_{35}H_{45}N_9O_{12}S + H]^+$ requires *m/z* = 815.2, found *m/z* = 815.3.

Synthesis of Methyl Sulfoxide Peptide, Ac-Asp-Phe-(MeS(O)Ala)-Ala-Thr-Nph-NH_2 (**18**)

The Ac-Asp-Phe-(S-methylcysteine)-Ala-Thr-Nph-NH_2 peptide (**39**) was synthesized on PAL-PEG resin following standard solid-phase protocols. After deprotection and cleavage from the

resin, the peptide was purified by RP-HPLC, lyophilized, and characterized by MALDI-TOF MS. The peptide (3.0 mg, 3.7 μmol) was then dissolved in MeOH:H_2O (1:1, 100 μL) and cooled to 0 °C, after which the oxidizing agent $NaIO_4$ (15.0 μmol) in H_2O (100 μL) was added dropwise. The solution was stirred for 1 hr, during which the formation of a white precipitate was observed. The suspension was warmed to room temperature and stirred for an additional 24 hrs. The reaction was then filtered, and the filtrate purified by HPLC to give the desired peptide (**18**) as a mixture of diastereomers. ESI-MS: calcd for $[C_{35}H_{46}N_8O_{13}S + H]^+$ requires *m/z* = 815.9, found *m/z* = 815.7.

Synthesis of Extended Asparagines (**12-14**) *and Aryl Asparagines* (**40-44**)

The Ac-Asp-Phe-Asp(OAl)-Ala-Thr-Nph-NH_2 peptide was synthesized on PAL-PEG resin (150 mg) using standard solid-phase protocols. Removal of the allyl group was carried out by incubation with $Pd(PPh_3)_4$ (20 mg) and phenylsilane (20 μL) in DMF (10 mL) for 30 mins while agitating with bubbling N_2. The solution was drained by vacuum, and the beads were washed (1 min) with DMF (5 x10 mL) and CH_2Cl_2 (3 x 10 mL); the procedure was repeated twice to give the free acid moiety. The beads were then taken up in DMF (10 mL), and PyAOP (62.5 mg, 0.18 mmol), 2,4,6-collidine (79.3 μL, 0.6 mmol), and the appropriate amine (0.06 mmol) were added in order. The suspension was mixed for 1 hr, the solution drained, and the beads washed with DMF/CH_2Cl_2 as previously described. The coupling was repeated twice. The peptide was cleaved from the resin following the standard procedure then purified by RP-HPLC to yield the desired peptides, which were characterized by ESI.

Synthesis of Neoglyconjugates (**45** and **46**)

The Ac-Asp-Phe-Asp(Hz)-Ala-Thr-Nph-NH2 peptide (**14**) was synthesized on PAL-PEG resin following standard solid-phase protocols. After deprotection and cleavage from the resin, the peptide was purified by RP-HPLC, lyophilized, and characterized by MALDI-TOF MS. The peptide (530 nmol) was then dissolved in DMSO (100 μL), and either GlcNAc (100 μL of 1 M solution dissolved in 1 M NaOAc, pH 5.6) or Bac was added. The pH was adjusted to pH 4.5 by the dropwise addition of glacial HOAc while stirring, and the reaction was allowed to stir at room temperature for 3 days while monitoring progress by HPLC. The glycopeptide was purified using by RP-HPLC gradient using a long gradient (5 min 5% B, 1 min transition to 20% B, 50 min transition to 60% B, 10 min transition to 95% B, 5 min 95% B, where B = CH_3CN + 0.1% TFA), and analyzed by ESI-MS.

Synthesis of Undeceprenylphosphate

Undecaprenylphosphate was prepared following previously published protocols.[49,50] Undecaprenol in hexanes (7 μmol) was placed in a round bottom flask and dried under vacuum for 3 hrs, after which tetrazole (0.027 mmol), anhydrous bis-(2-cyanoethyl)-*N,N*-diisopropylphosphoramidite (0.011 mmol), and freshly distilled THF (1.3 mL) were added. The reaction was stirred under argon for 2 hrs, then cooled to -78 °C for 10 mins. H_2O_2 (30%, 44 μL) was then added to a final concentration of 1%, and the mixture was allowed to stir for 15 mins before warming to room temperature over the course of 15 mins. The reaction was diluted with EtOAc (5 mL), quickly washed with 1% Na_2SO_3 (5 mL) and saturated NaCl (5 mL), dried over Na_2SO_4, then concentrated under vacuum. After solubilization in NaOMe/MeOH (1%, 5 mL) to remove the cyanoethyl protecting groups, the reaction was stirred for 48 hrs at room

temperature and followed by TLC. Undecaprenylphosphate was purified by DE-52 anion exchange chromatography, eluting with a stepwise gradient of NH_4OAc in MeOH (15 mM, 25 mM, 37.5 mM, 50 mM, 100 mM) to yield roughly 2.0 mg (~40%) of the desired product. ^{1}H and ^{31}P- NMR data matched previously reported values.

Synthesis of Und-PP-Bac-[^{3}H]GalNAc

The Und-PP-Bac-[^{3}H]GalNAc disaccharide donor was prepared as described previously.[51] Undecaprenol (10 μg, 13 nmol) was added to a 1.5 mL eppendorf tube and dried under a stream of nitrogen. After solubilization in DMSO (3 μL) followed by Triton X-100 (14.3%, 7 μL), 1 mM $MgCl_2$, 10 mM ATP, 1 M TrisOAc, pH 8.5, 50 μL *S. mutans* kinase CEF, 20 μL PglA, and 20 μL PglC CEF were combined in a final volume of 200 μL. A mixture of 2 mM UDP-Bac and UDP-[^{3}H]GalNAc (150 mCi/mmol, 5,250,000 DPM total) was prepared and added to the undecaprenol mixture to initiate the reaction. The reaction was incubated at 25 °C for 2 hrs, after which it was quenched by the addition of $CHCl_3$:MeOH (2:1, 1.3 mL). The organic layer was extracted with pure solvent upper phase (PSUP, 3 x 300 μL), then dried and stored at -20 °C until use. The overall reaction yield was determined by scintillation analysis of a small quantity of the reaction product, and was regularly found to be between 60-75%.

Cloning of PglB Mutants

Site-directed mutagenesis was performed by the polymerase chain reaction using a PglB pET24a(+) plasmid as the template, Pfu Turbo polymerase (Stratagene), and the oligonucleotides described in the Appendix (Table 1). The resulting PCR products were then

incubated with DpnI restriction endonuclease to remove the template and transformed into XL1 Blue competent cells. The final gene products encoded proteins with an N-terminal T7-tag and a C-terminal His_6-tag and contained the desired mutations.

Overexpression of PglB

The pET24a(+) plasmid containing PglB was transformed into *E. coli* BL21-CodonPlus(DE3) RIL competent cells (Stratagene) using both kanamycin (50 μg/mL) and chloramphenicol (30 μg/mL) for selection. For overexpression, 1 L of Luria-Bertani media supplemented with kanamycin and chloramphenicol was inoculated at 37 °C while shaking until an optical density (600 nm) of 0.6-0.8 was achieved. The cultures were then cooled to 16 °C and protein expression was induced by the addition of IPTG (1 mM). After 16 hrs, the cells were harvested by centrifugation (5000 x *g*) and the resultant cell pellets were stored at -80 °C until needed.

Preparation of PglB Cell Envelope Fraction

All steps were performed at 4 °C. A frozen cell pellet of *E. coli* expressing PglB was resuspended in 50 mM HEPES, pH 8.0 / 1 mM EDTA buffer (30 mL) supplemented with Protease Inhibitor Cocktail III (Calbiochem, 30 μL). The cells were lysed by sonication, and the mixture was centrifuged at 6000 x g for 30 mins to remove the cellular debris. The supernatant was gently decanted and subjected to a second round of centrifugation (37,000 x g) for 1 hr to collect the membrane fraction. The pellet was washed with 50 mM HEPES, pH 8.0 / 1 mM EDTA buffer, centrifuged (140,000 x g) for 1 hr, then homogenized in 50 mM HEPES, pH 8.0 /

100 mM NaCl buffer (3 mL). The crude cell membrane fraction was aliquoted and stored at -80 °C for future use.

PglB Activity Assay

To a tube of radiolabeled Und-PP-Bac-[^{3}H]GalNAc, DMSO (10 μL), 2x PglB Assay Buffer containing 100 mM HEPES, pH 7.5 / 280 mM sucrose / 2.4% v/v Triton X-100 (100 μL), 1 M $MnCl_2$ (2 μL), H_2O (53 μL), and 25 uL PglB freshly thawed cell envelope fraction were combined. The assay was initiated by the addition of 10 μL of a 2 mM stock of the peptide of interest in DMSO. Aliquots of the reaction mixture (25 μL) were removed every 2 mins and quenched by addition of 3:2:1 $CHCl_3$:MeOH: 4mM $MgCl_2$ (1.2 mL). The aqueous layer was then removed and the organic layer extracted with 600 uL theoretical upper phase (TUP) with salt (2 x 600 μL). The aqueous layers were combined, mixed with 5 mL of scintillation fluid (Ecolite, MP Biomedicals) and subjected to scintillation counting.

PglB Inhibition Assay

To test peptides as potential PglB inhibitors, the procedure outlined above was followed with slight modification; 10 μL of a DMSO stock of the peptide of interest was added to the Und-PP-Bac-[^{3}H]GalNAc substrate in the place of 10 μL DMSO. Upon addition of the remaining components, the entire mixture was allowed to incubate at room temperature for 20 minutes prior to reaction initiation with the acceptor peptide, Ac-Asp-Gln-Asn-Ala-Thr-Nph-NH_2 (10 uL of a 2 mM stock).

Screening PglB with Alkylating Agents

To a tube of radiolabeled Und-PP-Bac-[^{3}H]GalNAc, DMSO (10 μL), 2x PglB Assay Buffer containing 100 mM HEPES, pH 7.5 / 280 mM sucrose / 2.4% v/v Triton X-100 (100 μL), 1 M $MnCl_2$ (2 μL), H_2O (53 μL), and 25 uL PglB freshly thawed cell envelope fraction were combined. The appropriate alkylating agent was added to the reaction mixture to give a final concentration of 10 mM, and the reaction was incubated for 30 mins at room temperature. The mixture was then assayed for activity by adding the acceptor peptide Ac-Asp-Gln-Asn-Ala-Thr-Nph-NH_2 (10 uL of a 2 mM stock) and following the protocol outlined above for enzyme inactivation.

References

1. Weerapana, E.; Imperiali, B. Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **2006**, *16*, 91R-101R.
2. Nothaft, H.; Szymanski, C.M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **2010**, *8*, 765-778.
3. Yurist-Doutsch, S.; Chaban, B.; VanDyke, D.J.; Jarrell, K.F.; Eichler, J. Sweet to the extreme: protein glycosylation in archaea. *Mol. Microbiol.* **2008**, *68*, 1079-1084.
4. Imperiali, B.; Shannon, K.L.; Rickert, K.W. Role of peptide conformation in asparagine-linked glycosylation. *J. Am. Chem. Soc.* **1992**, *114*, 7942-7944.
5. Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **1983**, *209*, 331-336.
6. Marshall, R.D. Glycoproteins. *Annu. Rev. Biochem.* **1972**, *41*, 673-702.
7. Bause, E.; Legler, G. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-Glycosylation Step During Glycoprotein Biosynthesis. *Biochem. J.* **1981**, *195*, 639-644.
8. Imperiali, B.; Shannon, K.L.; Unno, M.; Rickert, K.W. A mechanistic proposal for asparagine-linked glycosylation. *J. Am. Chem. Soc.* **1992**, *114*, 7944-7945.
9. Imperiali, B.; Hendrickson, T.L. Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.* **1995**, *3*, 1565-1578.
10. Kellenberger, C.; Hendrickson, T.L.; Imperiali, B. Structural and functional analysis of peptidyl oligosaccharyl transferase inhibitors. *Biochemistry* **1997**, *36*, 12554-12559.
11. Weerapana, E.; Imperiali, B. Peptides to peptidomimetics: towards the design and synthesis of bioavailable inhibitors of oligosaccharyl transferase. *Org. Biomol. Chem.* **2003**, *1*, 93-99.
12. Peluso, S.; Ufret, M.d.L.; O'Reilly, M.K.; Imperiali, B. Neoglycopeptides as inhibitors of oligosaccharyl transferase: insight into negotiating product inhibition. *Chem. Biol.* **2002**, *9*, 1323-1328.
13. Tai, V.W.F.; Imperiali, B. Substrate specificity of the glycosyl donor for oligosaccharyl transferase. *J. Org. Chem.* **2001**, *66*, 6217-6228.
14. Wacker, M.; Feldman, M.F.; Callewaert, N.; Kowarik, M.; Clarke, B.R.; Pohl, N.; Hernandez, M.; Vines, E.D.; Valvano, M.; Whitfield, C.; Aebi, M. Substrate specificity of bacterial oligotransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7088-7903.
15. Szymanski, C.M.; Wren, B.W. Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **2005**, *3*, 225-237.
16. Blaser, M.J. Epidemiological and clinical features of *Campylobacter jejuni* infections. *J. Infect. Dis.* **1997**, *176*, S103-S105.
17. van Putten, J.P.M.; van Alphen, L.B.; Wösten, M.M.S.M.; de Zoete, M.R. Molecular mechanisms of *Campylobacter* infection. *Curr Top Microbiol Immunol* **2009**, *337*, 197-229.
18. Szymanski, C.M.; Burr, D.H.; Guerry, P. *Campylobacter* protein glycosylation affects host cell interactions. *Infect. Immun.* **2002**, *70*, 2242-2244.
19. Glover, K.J.; Weerapana, E.; Numao, S.; Imperiali, B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **2005**, *12*, 1311-1315.

20. Wacker, M.; Linton, D.; Hitchen, P.G.; Nita-Lazar, M.; Haslam, S.M.; North, S.J.; Panico, M.; Morris, H.R.; Dell, A.; Wren, B.W.; Aebi, M. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **2002**, *298*, 1790-1793.
21. Chen, M.M.; Glover, K.J.; Imperiali, B. From peptide to protein: comparative analysis of the substrate specificity of *N*-linked glycosylation in *C. jejuni*. *Biochemistry* **2007**, *46*, 5579-5585.
22. Sharma, C.B.; Lehle, L.; Tanner, W. N-Glycosylation of yeast proteins. Characterization of the solubilized oligosaccharyl transferase. *Eur. J. Biochem.* **1981**, *116*, 101-108.
23. Morrison, J.F.; Walsh, C.T. *The Behavior and Significance of Slow-Binding Enzyme Inhibitors*; John Wiley & Sons, Inc.: Hoboken, NJ, 2006; Vol. 61.
24. Ramsamy, K.; Olsen, R.K.; Emery, T. Synthesisof *N*-*t*-Boc-L-α-aminoadipic acid 1-*t*-butyl 6-ethyl ester from L-aspartic acid: a new route to L-α-aminoadipic acid. *Synthesis* **1982**, 42-43.
25. Werner, R.M.; Shokek, O.; Davis, J.T. Preparation of 4-Oxo-L-norvaline via diazomethane homologation of β-aspartyl semialdehyde. *J. Org. Chem.* **1997**, *62*, 8243-8246.
26. Peluso, S.; Imperiali, B. Asparagine surrogates for the assembly of N-linked glycosylation mimetics by chemoselective ligation. *Tetrahedron Lett.* **2001**, *42*, 2085-2087.
27. Spetzler, J.C.; Hoeg-Jensen, T. Preparation and application of O-amino-serine, Ams, a new building block in chemoselective ligation chemistry. *J Peptide Sci* **1999**, *5*, 582-592.
28. Bordwell, F.G. Equilibrium acidities in dimethyl sulfoxide solution. *Acc. Chem. Res* **1988**, *21*, 456-463.
29. Walter, W.; Becker, R.F. *Justus Liebigs Ann. Chem.* **1969**, *727*, 71-80.
30. Carrasco, M.; Jones, R.J.; Kamel, S.; Rapoport, H.; Truong, T. *N*-(Benzyloxycarbonyl)-L-vinylglycine methyl ester. *Organic Synthesis* **1992**, *70*, 29-34.
31. Tong, X.; Werner, R.M.; Lee, K.-C.; Fettinger, J.C.; Davis, J.T.; Coward, J.K. Synthesis and evaluation of tripeptides containing asparagine analogues as potential substrates or inhibitors of oligosaccharyltransferase. *J. Org. Chem.* **1998**, *63*, 4767-4778.
32. Laue, K.W.; Mück-Lichtenfeld, C.; Haufe, G. Enantioselective syntheses of 2-amino-4-fluoropent-4-enoic acids. Isosteres of asparagine. *Tetrahedron* **1999**, *55*, 10413-10424.
33. Takatsuki, A.; Arima, K.; Tamura, G. Tunicamycin, a new anitbiotic. I. Isolation and characterization of tunicamycin *J. Anitibiot.* **1971**, *24*, 215-223.
34. Varki, A.; Cummings, R.D.; Freeze, H.H.; Stanley, P.; Bertozzi, C.R.; Hart, G.W.; Etzler, M.E. *Essentials of Glycobiology*; 2nd Ed.; Cold Spring Harbor Laboratory Press: New York, 2008.
35. Wiesmann, C.; Hengstenberg, W.; Schulz, G.E. Crystal structure and mechanism of 6-phospho-β-galactosidase from *Lactococcus lactis*. *J. Mol. Biol.* **1997**, *269*, 851-860.
36. Johnson, P.E.; Tomme, P.; Joshi, M.D.; McIntosh, L.P. Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. *Biochemistry* **1996**, *35*, 13895-13906.
37. Van Liempt, E.; Imberty, A.; Bank, C.M.C.; Van Vliet, S.J.; Van Kooyk, Y.; Geijtenbeek, T.B.H.; Van Die, I. Molecular basis of the differences in binding properties

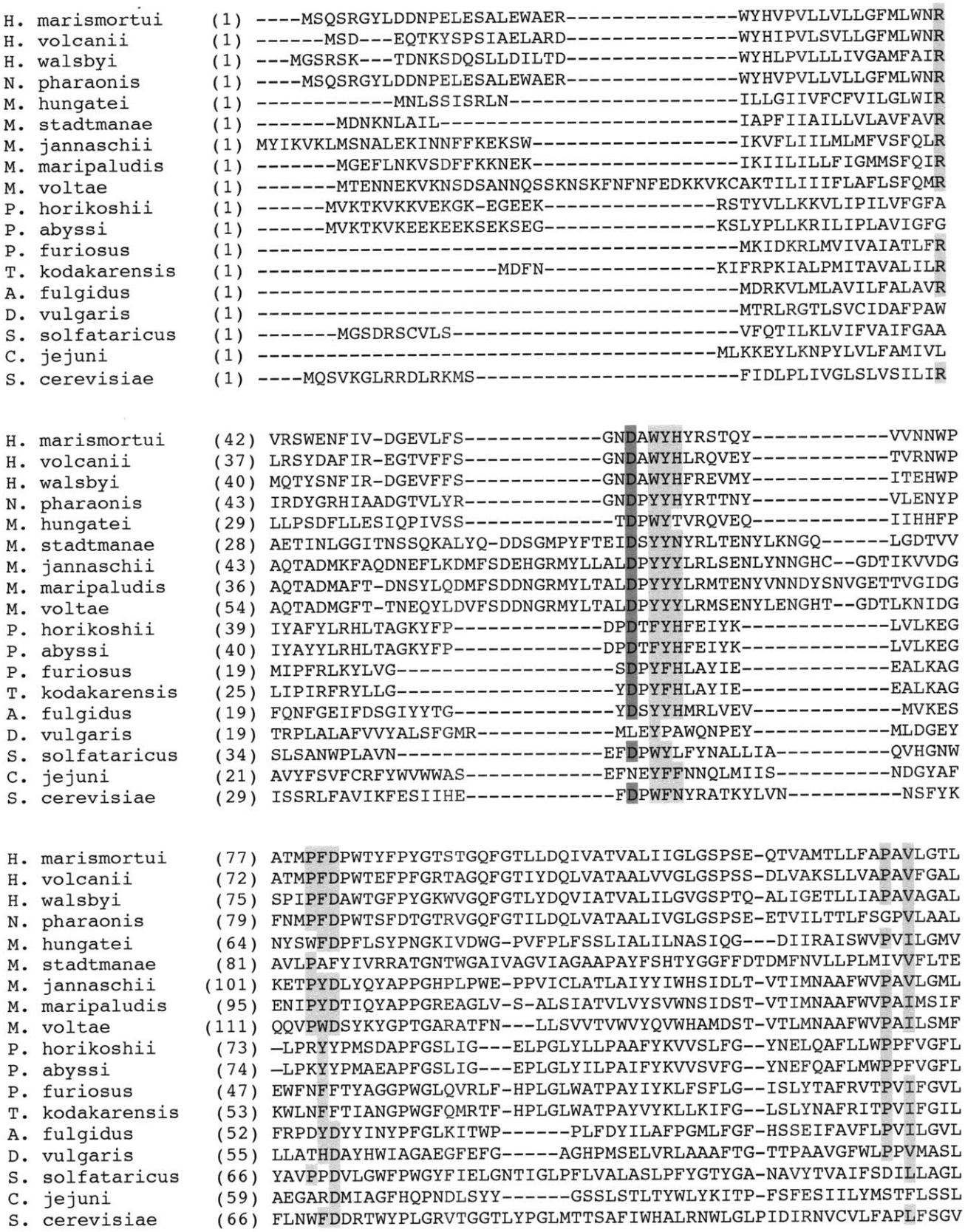
of the highly related C-type lectins DC-SIGN and L-SIGN to lewis X trisaccharide and *Schistosoma mansoni* egg antigens. *J. Biol. Chem.* **2004**, *279*, 33161-33167.
38. Pathak, R.; Hendrickson, T.L.; Imperiali, B. Sulfhydryl modification of the yeast Wbp1p inhibits oligosaccharyl transferase activity. *Biochemistry* **1995**, *34*, 4179-4185.
39. Miles, E.W. Modification of histidyl residues in proteins by diethylpyrocarbonate. *Methods in Enzymology* **1977**, *47*, 431-442.
40. Carrillo, N.; Vallejos, R.H. Essential histidyl residues of ferredoxin-$NADP^+$ oxidoreductase revealed by diethyl pyrocarbonate inactivation. *Biochemistry* **1983**, *22*, 5889-5897.
41. Abdulwajid, A.W.; Wu, F.Y.-H. Chemical modification of *Escherichia coli* RNA polymerase by diethyl pyrocarbonate: evidence of histidine requirement for enzyme activity and intrinsic zinc binding. *Biochemistry* **1986**, *25*, 8167-8172.
42. Glover, K.J.; Imperiali, B. Unpublished work, 2006.
43. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes *J. Mol. Biol.* **2001**, *305*, 567-580.
44. Igura, M.; Maita, N.; Kamishikiryo, J.; Yamada, M.; Obita, T.; Maenaka, K.; Kohda, D. Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **2008**, *27*, 234-243.
45. Maita, N.; Nyirenda, J.; Igura, M.; Kamishikiryo, J.; Kohda, D. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J. Biol. Chem.* **2010**, *285*, 4941-4950.
46. Nothaft, H.; Liu, X.; McNally, D.J.; Li, J.; Szymanski, C.M. Study of free oligosaccharides derived from the bacterial N-glycosylation pathway. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15019-15024.
47. Hancock, W.S.; Battersby, J.E. A new micro-test for the detection of incomplete coupling reactions in solid-phase peptide synthesis using 2,4,6-trinitrobenzene-sulphonic acid. *Anal. Biochem.* **1976**, *71*, 260-264.
48. Sanderson, J.M.; Singh, P.; Fishwick, C.W.G.; Findlay, J.B.C. The synthesis and reactivity of optically pure amino acids bearing side-chain thioamides. *J. Chem. Soc., Perkin Trans. 1* **2000**, 3227-3231.
49. Weerapana, E.; Glover, K.J.; Chen, M.M.; Imperiali, B. Investigating bacterial *N*-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **2005**, *127*, 13766-13767.
50. Ye, X.-Y.; Lo, M.-C.; Brunner, L.; Walker, D.; Kahne, D.; Walker, S. Better substrates for bacterial transglycosylases. *J. Am. Chem. Soc.* **2001**, *123*, 3155-3156.
51. Hartley, M.D.; Larkin, A.; Imperiali, B. Chemoenzymatic synthesis of polyprenol phosphates. *Bioorg. Med. Chem.* **2008**, *16*, 5149-5156.

Appendix

Table 1: Constructs and Oligonucleotides

Target Gene	Primer ID	Primer (5' to 3')
M. jannaschii stt3	F001	CGATGGATCCATGTATATAAAGGTGAAACTTATG
	R001	CGATCTCGAGTTTTAGATAATCAGTTCCATAATCTAC
M. voltae stt3	F002	CGATGGATCCATGACTGAAAACAACGAAAAAGTC
	R002	CGATCTCGAGTTTTGAGTAATTACCGTAATCCACG
A. fulgidus stt3	F003	CGATGGATCCATGGACAGAAAGGTATTGATGCTGGC
	R003	CGATCTCGAGGCTACCCACAACCTCAAAAATC
M. maripaludis stt3	F004	CGATGGATCCATGGGTGAATTTTTAAATAAAGTCTC
	R004	CGATCTCGAGATTGAGATAGTCAGTTCCATAATC
H. marismortui stt3	F005	CGATGGATCCATGAGTCAATCACGGGGCTATC
	R005	CGATAAGCTTTGCGCCCGCTTCAGCTGCGTC
wbpA	F013	CGATGGATCCATGATAGATGTTAACACAGTG
	R013	CGATCTCGAGAGCCTTGATGATGTGTGC
wbpB	F014	CGATGGATCCATGAAAAATTTCGCTCTCATC
	R014	CGATCTCGAGACGCGCAAGCGCCGCG
wbpE	F016	CGATGGATCCATGATTGAATTCATCGACC
	R016	CGATCTCGAGATTCGTCAACGCAGCACAG
wbpD	F015	CGATGGATCCATGAGTTATTATCAGCACC
	R015	CGATCTCGAGCACGTCCACCTTGCTCAG
pglB-H70A	F021.2	CACTTATTTTTGCTAGAAAAGAAAAGATTTTTTATATAGCTG
	R021.2	CAGCTATATAAAAAATCTTTTCTTTTCTAGCAAAAATAAGTG
pglB-H217A	F022.2	GTATGGCTTTTGAGAAAAGCTAAAAGTATGATTATGGCTTTACC
	R022.2	GGTAAAGCCATAATCATACTTTTAGCTTTTCTCAAAAGCCATAC
pglB-H350A	F023.2	GGCTCCAGTATTTATCGCTATTTACAACTATAAAGCGCC
	R023.2	GGCGCTTTATAGTTGTAAATAGCGATAAATACTGGAGCC
pglB-H423A	F025.2	CTTTTATGCTCCGCAAAATGCTATTTTAAAATCAGACATTTTACAAGC
	R025.2	GCTTGTAAAATGTCTGATTTTAAAATAGCATTTTGCGGAGCATAAAAG
pglB-D519A	F026.2	CTTTTATGCTCCGCAAAATGATATTTTAGCATCAGACATTTTACAAGC
	R026.2	GCTTGTAAAATGTCTGATGCTAAAATATCATTTTGCGGAGCATAAAAG
pglB-K522A	F027	GTAGATGGTGGAAAGGCTTTAGGTAAGGATAATTTTTTCCC
	R027	GGGAAAAAATTATCCTTACCTAAAGCCTTTCCACCATCTAC

Figure 1: Alignment of archaeal OTases with PglB (*C. jejuni*) and Stt3 (*S. cerevisiae*)



```
H. marismortui  (136) VAVPTYLVGRRLGGRLGGVTAVTVMAFSTGGLLQRSRVGFSDHHVAEALFQILGVLGVMV
H. volcanii     (131) TVIPTYLIGKRLGGRLGGLFGAVILMLLPGTFLQRGLVGFADHNIVEPFFMGFAVLAIMI
H. walsbyi      (134) AVIPVFFIIKSLSGRIAGIFGAVVLMLLSGTFLNRTLVGVADHNAIEPLMLGIAVFGLVV
N. pharaonis    (138) CAIPLYFIGKRLGGRFGGIIAVVVLALTAGQFLTRSVAGYYGHHVTEVLFTLIALLVGMK
M. hungatei     (120) LIPLCYMFGKIIWDVKAGWFAAILICVVGGETLFRSFFGDVDHHIMEVVVTSSFFISYFI
M. stadtmanae   (137) AVLPAFYIVRRATGNTWGAIVAGVIAGAAPAYFSHTYGGFFDTDMFNVLLPLMIVVFLTE
M. jannaschii   (159) LGIPIYFVVRRVTNSNIGGIAGAIALISAPGLLYKTCAGFADTPIFEVLPILFIVWFILE
M. maripaludis  (152) LGIPVFFIVRRNTASNIGGLVGALLLISSPSLLYKTSAGFSDTPIFEILPLLFIVWMIME
M. voltae       (167) LITPIFFTVRRITSSDIGGAVAAILASLSPSIFVKTVAGFSDTPILEILPLLFIVWFIIE
P. horikoshii   (127) GVIAVYLLGRKVLNEWT-GLWGAVVLTVSTANFSRTFSGNARGDGPFMALFIFASVAMLY
P. abyssi       (128) GVIAVYLLGRKVLNEWA-GLWAAVILSVSTANFSRTFSGNARGDGPFMTLFLFSLVAMLY
P. furiosus     (104) TVVFFYLSLKKLYNRDV-AFIVGLFLGVNYGHIFRSMANYYRGDNYMLFWYSVALLGIAL
T. kodakarensis (110) TIAVFYWALLKLYDEKR-AFFASLFLAVSFGHVFRSMAGYYRGDNYMLFWYSISLLGIAY
A. fulgidus     (105) SVVLICLTALQIVNNQTFALISAFIYAAAPVAVWKTVLGQADHHALVIFLFLLSAYLLLK
D. vulgaris     (109) VAVGVFGWAAALGGVEAGVCAG-VLASLAPGFLARTLLGYCDTDLVTLLFPLLMGLGPAF
S. solfataricus (125) GVIASYLSIESITNSRLAGYMAAAIIAVSPALTYKNLLGGLPKTSWGAVFILFAIFLLNQ
C. jejuni       (112) VVIPIILLAN-EYKRPLMGFVAALLASVANSYYNRTMSGYYDTDMLVIVLPMFILFFMVR
S. cerevisiae   (126) TAWATYEFTKEIKDASAGLLAAGFIAIVPG-YISRSVAGSYDNEAIAITLLMVTFMFWIK

H. marismortui  (196) AVSVAARDKPIYEQFIER----------------------------------------
H. volcanii     (191) ALTVADREKPVWELVAAR----------------------------------------
H. walsbyi      (194) AFQKAEATMPVWEVVYEEVFD-------------------------------------
N. pharaonis    (198) MVAVAQREKPIYEFVRTR----------------------------------------
M. hungatei     (180) LLSYFN----------------------------------------------------
M. stadtmanae   (197) SVYAS-----------------------------------------------------
M. jannaschii   (219) SIHSQEKTALFKKDLKNPISLFVIAALIIELIIGAYLNIASGESVVIASILFYTVSLAFI
M. maripaludis  (212) AIHE----------------------------------------------------
M. voltae       (227) AIHYS---------------------------------------------------
P. horikoshii   (186) YLKES---------------------------------------------------
P. abyssi       (187) YLKEN---------------------------------------------------
P. furiosus     (163) GLKTR---------------------------------------------------
T. kodakarensis (169) AFSTR---------------------------------------------------
A. fulgidus     (165) ----------------------------------------------------------
D. vulgaris     (168) WMARFMFTPLQVAR--------------------------------------------
S. solfataricus (185) GLKKK---------------------------------------------------
C. jejuni       (171) MILKK---------------------------------------------------
S. cerevisiae   (185) AQKTG---------------------------------------------------

H. marismortui  (214) ----------DIDALRDTISWSMLAGVAIAMYLWVWPPGVLLLGILGVFFLLRLCLEYV
H. volcanii     (209) ----------DLDALREPLKWSVLAGVATAIYMWSWPPGILLVGIFGLFLVLKMASDYV
H. walsbyi      (215) ---------NHDIDSIEEPLKWGLIGGVLTGLYLWTWPPGVFIVGIVGIFTILKISSDVV
N. pharaonis    (216) ----------EFDLLREPVLWGAAFGVTLVVAILNWPPAVFLFGVFAAFLFVQLSLEFI
M. hungatei     (186) ----------ANNNNRTKWKAPLSTKSWSHHYFGNSKIQNLVIVLSIICGILYYLAI
M. stadtmanae   (202) ----------KPLFKGLYAGIASLFLGLFALSWSGYTYMVLLTFGTLILFIPISYI
M. jannaschii   (279) LAGLIIAGIKKLKGNELEFELFALLAVILTAVSPKMWGAWWYGFDVITAFLVIYIIALAL
M. maripaludis  (216) ----------QENSKKSGIFGGIAAILIGLYPMMWSGWWYAFDITAGFLVLYTAYETL
M. voltae       (232) ----------KEKNYKSLIYGLLATLMLALYPFMWSAWWYGYYIVIAFLVIYAIYKGI
P. horikoshii   (191) ----------NKT--RKIIYGTLFVLLTVISLGAWNGSPFGLMVLLGFASLQTIILFI
P. abyssi       (192) ----------DIK--KKSLWGAVFVLLASISLGAWNGSPFGLMVLIGFASFQTIALFI
P. furiosus     (168) ----------SKY---RYLFYLLPGIATGFASAFWQAYYPIFVFVLAGGLLLGVTAYL
T. kodakarensis (174) ----------ERLGHKSLAFYAVPALASGLAAAFWQAYYPIFVFLLASAVFLGVGDFL
A. fulgidus     (165) ----------DGVWKILAGLPMLFMALAWLGSPIYGALLAFSALVH-----F
D. vulgaris     (182) ----------RLLARRTGTSDVVPDAVAVMMTSRGEADDDWHGGALSPRWLALLAGSGL
S. solfataricus (190) ----------NIWYGIPAGIVLFLAEISWGGYTYIDLSLLVAAFLLILLNRN
C. jejuni       (176) ----------DFFSLIALPLFIGIYLWWYPSSYTLNVALIGLFLIYTLIFHR
S. cerevisiae   (190) ----------SIMHATCAALFYFYMVSAWGGYVFITNLIPLHVFLLILMGRY
```

```
H. marismortui   (263) HDS----------------------------------SPEHTAIAGAITMVTAAVVSAS
H. volcanii      (258) RGR----------------------------------SPEHTAFVGAISMTVTGLLMFI
H. walsbyi       (266) NER----------------------------------TPEPTAFAVVVTMSVVAVMSLI
N. pharaonis     (265) RGH----------------------------------SPDHIAIPSVVAMVATAVLLLP
M. hungatei      (233) MTMP---------------------------------TCTLIAIIIALFTFFLPVLITE
M. stadtmanae    (248) MDK--------------------------KDGKDLSNKKQWLKNNPATLPLVVFIIL
M. jannaschii    (339) LKS----QVKIKEFINIG-------NLKNIVYLSIFYIFGSFVLLVAIYGMGIAISPITS
M. maripaludis   (264) --------TKSKN------------LKNVITTSLITLVGGAILVSLSTGLSGFINWILS
M. voltae        (280) SYNSIAKYTKSKNNNHKDKIESEKLEMLNILKISGLFIIGGAVLITALYGVSTTMNALQA
P. horikoshii    (237) FGK----------------------------------LEELKKFVKEFYPAYLAILAFGYAL
P. abyssi        (238) FGK----------------------------------IKELKKFVKEFYPAYLAILAIGYGL
P. furiosus      (213) KS-----------------------------------PKLFLDSILIVLSTGLGVLIANILG
T. kodakarensis  (222) LGR----------------------------------DRYILDGILLTVSTAIGALIANYLG
A. fulgidus      (202) DRK------------------------------------ALRLVAASYLIPAISFVLY
D. vulgaris      (231) VGSWG--------------------------AEWHSLFPYLIRYDVVLLFGCICVFG
S. solfataricus  (232) DE-------------------------------------------ITANLYTIMVVVTSFL
C. jejuni        (218) KEK-----------------------------------------IFYIAVILSSLTLSN-
S. cerevisiae    (232) SSK----------------------------------LYSAYTTWYAIGTVASMQIPFV

H. marismortui   (288) RINVLEITATAQSLLQPGLALAVGFGCVFMAWLARFMEQREYGQYVYPATVFGILAFGAV
H. volcanii      (283) PIEEPGFGVTDFGFLQPLFSLGVALGAVFLAALARWWESNDVDERYYPAVVGGTMLVGIV
H. walsbyi       (291) AIDRVQFDTTSLSLLQPVAAIGVAGSAIALSWLARIWERKNIDISLYPVAVTGLATVGIA
N. pharaonis     (290) FIQTTELTVTDYSVFHLAFALGVAAGAVVMAGVARLWERRDIARRAYPAGVVGVGLIVGA
M. hungatei      (259) KSDYSRIFVMNGIIFGLFIILFALSGIQVSGWSLGQYSIIHLFIPAGIIAESGILYILSR
M. stadtmanae    (279) SIIILALTVGSSMFESITTVLGSATSLQSATAGTAYPNVYVSVGELQIPQVIDVANQSGG
M. jannaschii    (388) PLGYNQILSTYTQTTGWPNVYTTVAELAKPSSWSEIFTNAIGSDTIAIVGILGILLSFLS
M. maripaludis   (303) PIGF-TVINEATKITGWPNVYMTVSELAIPT-VTDIIENSVGNIWLLIAGISGILLSFVS
M. voltae        (340) PLNY-LGLDEVSSQTGWPNVLTTVSELDTAS-LDEIISSSLGSIHLFAIGLIGIFLSLFR
P. horikoshii    (265) TFPGIVKIGGFIRFAFEVFLGLIFLLVIMLY-GGRYLNYSDKKHRFLVVTIIVLLGFGGA
P. abyssi        (266) TIPGIAKIGGFIKFAFEVFLGLVLLVTIMLY-GGKFLNYSDKKHRFAVVAVIVLLGFAGA
P. furiosus      (240) DVKGYGMLG-YTDWMGKKVAETFGLEFGFIK-DAYLLIHVKYLLPLSLV-FLGFLIITKK
T. kodakarensis  (250) ARVGYGMLG-YDRWLSKTVAEKLSLELTTVR-DAYLLFHLKYLVPLALAGLLALLALSKF
A. fulgidus      (224) PPVGISFFG---------LAAFLFVGSVVKGYEDRFRNATIYYIALSLATVLIIYFIPLP
D. vulgaris      (262) RPGMRRALLVGGLAYALPMLVGAWGLLFALALGWGLHRADARFMAFVRGRALPVAGWCAV
S. solfataricus  (250) TSLAPNNIGFMSGLAHGFSLLLISAVLYLDLYLSRTLPKEIVDSRNLIVIAVLIFILTLG
C. jejuni        (236) --------------------IAWFYQSAIIVILFALFALEQKRLNFMIIGILGSATLIFL
S. cerevisiae    (257) GFLPIRSNDHMAALGVFGLIQIVAFGDFVKGQISTAKFKVIMMVSLFLILVLGVVGLSAL

H. marismortui   (348) LMAVLTP-------DLWSFFTNNVMRIIGLTTSETANTVGEASPLRNAGVLFDR------
H. volcanii      (343) LFSLVLP-------SVFDSIARNFLRTVGFSAGAATRTISEAQPFLAANVLQSNGQTAVG
H. walsbyi       (351) VLSLLQI-------GLIDLITGNLLRIVGFSASATARTVGEAQPFVSQGSLRRFGVSIPG
N. pharaonis     (350) VIAVALP-------SVFDFLISQVERVAGFGATDTRATIGEAQAPDSPVSFFFAN-----
M. hungatei      (319) FIVAKNRFFYTGILLMIFFVFSALLILINSEISNSFTNMVTRFFALNGTSVTIQEMQSSN
M. stadtmanae    (338) LFSLLYA-------VAALFMLGYVLRKKPVNKDKKDETEEKTETIEDKKPRNRRYTPKTK
M. jannaschii    (448) LRYE----------KVKLDIKYSILLAIWLAVTLYAATKGIRFAALATPPLAIGLGIFVG
M. maripaludis   (361) FKHD----------KQKIDIKYALYLTLWLIATVYAATKGIRFVALMTPALAIGIGIFAG
M. voltae        (398) KVLTPVKQISN-GLAEKLDIKYALLLIIWFAVTFLAASKGVRFVALMVPPLSIGVGIFVG
P. horikoshii    (324) YAYVG---------PKLFRLMGG----AYQSTQVYETVQELAKTTIGDVKAYYGVESGNG
P. abyssi        (325) YAYVG---------PKLFRLMGG----AYQSTQVYQTVQELAKTTLSDIKLYYGVEGNNG
P. furiosus      (297) LNPK----------IKVGVLVGGSILAFIVMLVKFPALKDLSTGFGTFREVPISETLPPT
T. kodakarensis  (308) IRDKRL--------RTLTVLVFG-ALAVYLLFARFEGLKDLSTGFGIFNEWPISETRPST
A. fulgidus      (275) HFEFVKG------------GINYIFGANIYLPTISEARSLQIFEIISASGYIYFIFALIS
D. vulgaris      (322) AVLLFDP------DVFRMFRLNSIAGYVKRSGDPSPGAGGDDPLVYPSVAQSIIEVQDLS
S. solfataricus  (310) LSG-----------------LSILRPTSAPIPSRYYAIINPFYQVTVPIDKTVAEYIPQP
C. jejuni        (276) ILSGG-----------VDPILYQLKFYIFRSDESANLTQGFMYFNVNQTIQEVENVDFSE
S. cerevisiae    (317) TMYGLIAPWTG---RFYSLWDTNYAKIHIPIIASVSEHQPVSWPAFFFDTHFLIWLFPAG
```

```
H. marismortui   (348) -----HGFAVIVAGIGGLLLLGKQFLSDDAP----------------------------
H. volcanii      (343) RIMSEYGFTFFTGALAAVWLVAKPLVKGGNSRKIGYAVGSLALIGVLFLIPALPAGIGSA
H. walsbyi       (351) RIAVEYGLTFFTGLAAAFILHAKPLFKKGTQRAYAYLGVGISIIGLLFLAGFIPDALETV
N. pharaonis     (350) -----YGLGIYTALAGFGLLLYRVFSTKRPQ----------------------------
M. hungatei      (319) IFTFLKVFNITAILSCFGFLILCFQFYRTKN----------------------------
M. stadtmanae    (339) KAEEVMHHEDVFKPGRYELTSKEKYNNLFLY----------------------------
M. jannaschii    (448) QLERFLKMKSDIAIFGIGIP-AGIFGLLILS----------------------------
M. maripaludis   (361) QIENIIKR-YEKKVEYILYPVIGILSVITLI----------------------------
M. voltae        (398) FIEQFIKNNLDKKYEYVAYPTIAIIVLYALF----------------------------
P. horikoshii    (324) LIFFLSIPGLLILLTKYLYDLFKKAKSDNET----------------------------
P. abyssi        (325) LVFFLSIPGFLIILGLYLNALLKKSESSNEY----------------------------
P. furiosus      (297) LDDLWRAYNIAIFLAALYILRLR--KIRSGD----------------------------
T. kodakarensis  (308) FHDLWSAFAVGLFLAPLFFLRFRPGKVRVQD----------------------------
A. fulgidus      (275) VLFFRNRFVLSMFFLSFGFVWLLRKHKSMIM----------------------------
D. vulgaris      (322) LSEVLSYFHPWLWVALAGLGGFVPLLLARPA----------------------------
S. solfataricus  (310) ITAMIQDFGIALFLSVIGMYYLIRRGNLVGL----------------------------
C. jejuni        (276) FMRRISGSEIVFLFSLFGFVWLLRKHKSMIM----------------------------
S. cerevisiae    (317) VFLLFLDLKDEHVFVIAYSVLCSYFAGVMVR----------------------------

H. marismortui   (395) --------------------------AEEMLVVVWAVFILLATFTQRRFAYYLVAPIAV
H. volcanii      (396) LGVEPSLVSLTIVTALIVGAVMQADYESERLFVLVWAAIITSAAFTQVRFNYYLAVVVAV
H. walsbyi       (404) TGLDEQVAALLVVSAIIGGATFLASYDADKLFIIVWAAFITGMAFTQVRFNYYLAIVVAV
N. pharaonis     (398) --------------------------AEYLLIAVFSAFMLAFTLTQVRFDYYFVIAVGA
M. hungatei      (379) --------------------------PLNLAVIIWASVYIVISLLVVRFQYYSGEIFVI
M. stadtmanae    (392) ------------------------------VLTLLWILGIAIMLTQGTRFIEQFAVPVAL
M. jannaschii    (498) ------------------------------KYS----AKISQILLPTTYVPIIAYGFLIV
M. maripaludis   (411) ------------------------------KYS----GELFNILVPTTYVPIAVYLSIIA
M. voltae        (457) ------------------------------TIYRADSADLVRMLLPSNYVPIAEGIMLAS
P. horikoshii    (371) ------------------------------LFALVFYTMSLYLLYLAVRFLFLASYAVAL
P. abyssi        (372) ------------------------------MLSLVFYIMSLYLLSLAVRFLFLASYAIAL
P. furiosus      (347) ------------------------------AILLGYVITSLWMLRYWTRFLFLASYAIAL
T. kodakarensis  (359) ------------------------------FMLLGLVLPSLYMIKTWTRFLFIGSMGVAL
A. fulgidus      (323) ------------------------------VPSALLSAYLVSLVLERLEYPVFEKADEEE
D. vulgaris      (375) ------------------------------AFFLLPLGVLSFLSVKLGGRMVMFGAPVLAL
S. solfataricus  (353) ------------------------------WLLVLGVASIFGTSEQPYLFNYTAYMVAT
C. jejuni        (325) ------------------------------ALPILVLGFLALKGGLRFTIYSVPVMAL
S. cerevisiae    (374) ------------------------------LMLTLTPVICVSAAVALSKIFDIYLDFKT

H. marismortui   (421) LSGMVVGRFIRWFDFSAD-----DGIEYYQVMTILTVLLVIVVPLFAFGTTP--------
H. volcanii      (456) MNAYLLREALGIDFVGLANVERFDDISYGQVAAVVIAVLLILTPVLIIPIQLGNGGVSQT
H. walsbyi       (464) FNGYLFGQLLGYLDLKRSLTSLKNDIDGYQVLAIGAAILLILGPGLAVPISLGNTTTSPA
N. pharaonis     (424) GNAYLVGWVYQYVDLDNVR-QDFTNIQPYQILIVIAILFVIAGVPLVTGATLA-------
M. hungatei      (410) LAGVFLSALYDKIQVRHHKIQKNDKNKFFWKIIFINNKGILCIGFLITMITILSIPVAME
M. stadtmanae    (423) LSGLFVGIMVKYIDLKTENYSYVALIAAILVLLAVVSPLYADHLASS-------------
M. jannaschii    (528) LALLAIYKISDIISTLNDKKETIIKVSTLLLCIGVVIPPLSAVVPFS-------------
M. maripaludis   (441) LLVLAVYKIIDIISE---KEQAVKKVFGILLAFMLVFPSMAAAVPFY-------------
M. voltae        (488) LAVLIIYKVAELIAESN-KKLVMNKIFMILLAIGLITPTIATIVPFY-------------
P. horikoshii    (402) FFGIFIGFSMDVIEKMKENIGIKAALGIVLSLMILVIPFVHAPVLARSARALK-------
P. abyssi        (403) FSGIFAGFTMEVIEKMKENVGIKAALGIAIAVMILMVPITHGPVIARNAKALK-------
P. furiosus      (376) LSGIG---VYELTRRIKENKIRITSLGVVILLS---SAFSLGEVYS-------------
T. kodakarensis  (390) MAGIGLLELYEILPRFEGRKGLAVALGLLLFVPAVDTAIGFEKVAA-------------
A. fulgidus      (354) KSRRRKRKDRKVKQKNAEVEWKDHAVVAAFLVILAIPCIVVAVVPFD-------------
D. vulgaris      (406) GFALPMVWGVQRVLRHDLRTGWVRLALSAVLLVIVATPFVDLLPAMT-------------
S. solfataricus  (384) LGGLGVYYISDNLLKGVKNGNGNKILVGFILALVGISLVADAGLATLASNEPP-------
C. jejuni        (356) GFGFLLSEFKAILVKKYSQLTSNVCIVFATILTLAPVFIHIYNYKAP-------------
S. cerevisiae    (405) SDRKYAIKPAALLAKLIVSGSFIFYLYLFVFHSTWVTRTAYSSPSVVLP-----------
```

```
H. marismortui   (454) -VEAGSTGPGPSIQGWDESLQWMEGNTPAEGAYGSGGDAT-LEHYGTYERQEDFDYQEGQ
H. volcanii      (516) AMQASQTGPG-TVTQWDGSLTWMQNNTPAEGEFG--GESNRMEYYGTYEYTDDFDYPDGA
H. walsbyi       (524) WEAAQNNGPG-AVTVWDDSLEWMQGNTPEEGNLGGAGNADEMELYGTYEFTDDFEYPAGT
N. pharaonis     (457) -TADDASQPG-EMEQWTDSLDWLSEETPEPGAYG-TGDDPRLDYYGTYEPTDDFEYEAGE
M. hungatei        (1) VGLNETPKDS-INDEWIQSLKWLSQQNNTDVP-----------DYYSIYKESFSYPQNI
M. stadtmanae    (453) -QAVGSTNDD-----MYNTLTWIKANTSQD-------------------------------
M. jannaschii    (554) --VAPTFNNG-----WKEGLDWIKANTPNN-------------------------------
M. maripaludis   (467) --TAPTMNNG-----WMDSLSWIKSETPEN-------------------------------
M. voltae        (517) --SVPTYNDG-----WEGSLEWINTQTPNN-------------------------------
P. horikoshii    (432) --NTEIEVTG-----WEQALKWLRSNTSKY-------------------------------
P. abyssi        (433) --VSEIETTG-----WEQVLKWLNENTSKY-------------------------------
P. furiosus      (406) --VKPFMNEN-----WEKALIFIRENSNEN-------------------------------
T. kodakarensis  (420) --VEPFMNEH-----WERALTWLGENSNEN-------------------------------
A. fulgidus      (384) ------LTED-----WKEALEWMRTSLEEQNYLN-------------------PYEKPE
D. vulgaris      (437) --QGPIINKR-----HAEALRHIRTATPED----------------------------
S. solfataricus  (413) AITNAATSFLTTNYSWVSAANWIRTHSPQN-----------------------------
C. jejuni        (384) ---TVFSQNE------ASLLNQLKNIANRE-----------------------------
S. cerevisiae    (434) --SQTPDGKLALIDDFREAYYWLRMNSDED-----------------------------

H. marismortui   (501) YGVLSWWDYGHWIT--TRAERVPNANP-FQQGTETVAPFLVAQNETQANNILDST--DED
H. volcanii      (576) YGYMSWWDYGHWIT--VLGERIPNANP-FQGGATEAANYLLAEDEQQAESLVTSMGDDGE
H. walsbyi       (584) YGVQSWWDYGHWIT--TQSNRIPNANP-FQQGATSAANFLLSPNETQSHTVLNEQ--STE
N. pharaonis     (509) YGVLSWWDYGHYIT--TRGERIPVANP-FQHHATESADFLLADDEDEALDLLDEN--HGD
M. hungatei      (503) STVISWWDYGHWIL--VLAHKIPATSP-FQDNIVPVAKFLLADSNIQAEMIASQL-----
M. stadtmanae    (500) TVLASWWDFGHLFT--AVADRQVVFDG---GSQNNMRAYWIGNALTSTDEAKSAG-----
M. jannaschii    (601) SVITCWWDNGHIYT--YEARRMVTFDG---GSQNSPRAYWVGRAFATSNENLSIG-----
M. maripaludis   (511) SVVTCWWDNGHIYT--WATRKMVTFDG---GSQNTPRAYWVGHAFSTSDENLSVG-----
M. voltae        (564) SVVTCWWDNGHITY--WKTDRMVTFDG---SSQNTPRAYWVGRAFSTSNESLANG-----
P. horikoshii    (485) ATATSWWDYGYWIESSLLGNRRASADGGHARDRDHILALFLARDGNISEVDFESW-----
P. abyssi        (486) ATATSWWDYGYWIESSLLGHRRASADGGHARDRDHILALFLARDGNVSEVDFESW-----
P. furiosus      (446) DIVLTWWDWGHFVT--YYARRSPVAQG---SPNSGVAGYYLGLVDNGWAQSLG-------
T. kodakarensis  (466) DIVLAWWDYGTWVT--YYARRAPVAEL---APNTGVALYYLGKRDENWAMGLG-------
A. fulgidus      (431) YSVMSWWDYGNWIL--YVSKKAVVCNN-FQAGAVDAAKFFTAKSEDEAIKIAKKR-----
D. vulgaris      (484) SMVWIWWDWGYSAH--HFAHRRTIADG---ASHGGPSLYVPAAVFSTANPRFAWQ-----
S. solfataricus  (466) AFVLSWWDYGYWLE-VLTNRSVIDENNTLNGTQIRLMAEMFLNNETFAANVLENDFHLYP
C. jejuni        (431) DYVVTWWDYGYPVR-YYSDVKTLVDGGKHLGKDNFFPSFSLSKDEQAAANMARLS-----
S. cerevisiae    (483) SKVAAWWDYGYQIG-GMADRTTLVDNNTWNNTHIAIVGKAMASPEEKSYEILKEH-----

H. marismortui   (559) DAKTRYIAVDWKMAETNGNLNGKFFAPATFAD---GVEPGDYYSRMYIQSQDGAR-----
H. volcanii      (633) GDQTRYVMVDWQMASTD----AFKSAPTVFYD-ESNISRSDFYNPMFRLQEQGEQTTVAA
H. walsbyi       (643) GNQTRYVMVDWQMATPG----SKFGAPTVFYNAEPNVSRSDFLRTVYQFNQEGR---FSG
N. pharaonis     (566) GEGVQYVMVDYQLGYAG---TQKYGAPTAFES-EHDISDGDVGIQVIN-PETGEF-----
M. hungatei      (550) --HGKYIITTNEYLFNDFPMIQKWIPAQPDEDPYYFVFYNRASKTSSILTPMLG------
M. stadtmanae    (524) --ILRMLANSGEDASNT----LDLYTNNTEKTVEILNAILPMDRTEANSALTGT------
M. jannaschii    (624) --IIRMLATSGDEAFKKGSVLMNFTHNNVSKTVKILNEILPVDRSKAYDILTKK------
M. maripaludis   (534) --ILRMLATSGDSAYDDDSILIKKTG-SIKDTVDILNKILPLTRTEAKASLVNN------
M. voltae        (587) --IFRMLASSGDKAYTTDSVLIKKTG-SIKNTVDVLNEILPLTKSDAQKALKNSS-----
P. horikoshii    (508) --ELNYFIIYLNDWAKFN--AISYLGGAITRKEYNGDENGRGRVTTILLTQAAG------
P. abyssi        (509) --ELNYFIIYLNDWAKFN--AISYLGGAITRREYNGDETGRGQVTTILPLQGSG------
P. furiosus      (469) ---VDYVIVSLYDILKFE--AIVDTAKLSRKWENISRADYG---VDFLKLTEST------
T. kodakarensis  (489) ---VDYVIVSYYDFLKFG--TIVDTAMMCSRCN--VSENYG---LVVLPMVSSA------
A. fulgidus      (460) --GVRYVVTADEITMKDAN-----------------------NTKFPAIMRIAG-----
D. vulgaris      (507) --LIRYTAERGG-------------------------------IPGSVFEGMGG------
S. solfataricus  (496) YGSPNYTIPVYIVAYDAVTFVYAGSQAQWYVGYPPSFPGAFFGYTTSLGDIGKAMG----
C. jejuni        (452) ---VEYTEKSFYAPQND-------------------ILKSDILQAMMKDYNQSN------
S. cerevisiae    (511) --DVDYVLVIFGGLIGFG----------------------GDDINKFLWMIRIS------
```

```
H. marismortui   (614) QITFQQQAYYETMTARLYRFHGSAAEPQPIVVDWENKQTGQGAQYRGAPTNEQEPEGQRQ
H. volcanii      (690) ASSLKDQRYYESLMVRLYAYHGSAREASPIVVDWEER-----TSADGSTTFRVTPSDGQ-
H. walsbyi       (698) TTSVRTDRFYNSTMTRLYYYHGSARQPSPIVVDWENRRV--QTNDGGDITVAANPRGNQS
N. pharaonis     (621) VYGAHTQRGYDSMRVQLYQHHGSAQEPASFTLQLGEY--------DEASGIATIPEGAQP
M. hungatei      (602) ----MKPAFFNTTLVKLHANDGSYVHANGSVVIQYQSTIVSGKEVPLLQKMYPIDQTQTI
M. stadtmanae    (574) -----YGLSQQEANSVLDLTHPAKVKPVNLILSSDMLSKAAWWSYFGSWDFKNQNSTHYS
M. jannaschii    (674) -----YGLSDKKAKLVLNATHPEHPNPDYLITYNRMTDIAPVWSMFGFWNFSLPPNTPND
M. maripaludis   (584) -----YDLTDAEAEEVLDLTHPKVTNPDYLITYNRMTSIASVWSMFGNWNFSLPASTENS
M. voltae        (637) -----YKFTDTEVSEILDATHPKVTNPDYLITYNRMTSIASVWSYFGNWDFNLPAGTSRS
P. horikoshii    (563) ------NVYVNPYARIVIKVIQQNKTRRIAVNIG-QLECSPILSVAFPGNIKIKGSGRCS
P. abyssi        (564) ------GIYVNPYAGISVRVVQSNTTSKVTVNVRGRAECSPIYTLLIPGNKKIPGNGRCS
P. furiosus      (517) ------GSILRFDSQYSTLIVKEG---NIRVILSGKVVY-PREAIIES-NGRIKNLKYPA
T. kodakarensis  (537) ------GALVFRNSGY-TVVARPGEKWDVKISVSGHVFG-PREVYVEYGDKIIKPNITPS
A. fulgidus      (512) -----YNVDLMTEGEILNFFNHTVLYRLHMENAENLTHFRLVKEFGDVKIFEVVGS----
D. vulgaris      (557) -----AQAQELVHRLGIEKMTFADAPRQYLVVSYDMLRLGFWITNFGTWDFLSREGKGYA
S. solfataricus  (555) -----AMTTIAGYPLDEYVNLTEINNTITQIINTYATTNPTIAQTLASQVSQAEPFAWTP
C. jejuni        (506) -----VDLFLASLSKPDFKIDTPKTRDIYLYMPARMSLIFSTVASFSFINLDTGVLDKPF
S. cerevisiae    (565) -----EGIWPEEIKERDFYTAEGEYRVDARASETMRNSLLYKMSYKDFPQLFNGGQATDR

H. marismortui   (666) GQIVRQFESMEQARQYVEEDATAQIG----------------------------GFGDN
H. volcanii      (745) --AVRTFDNMSAAEEYVANDPTSQIG----------------------------GIGTF
H. walsbyi       (751) --TVRTFNSMREAEAFVANDTTSQLG----------------------------GFGAY
N. pharaonis     (671) DDLIQEHASPEEAREAAQAPNVIHGG----------------------------VLGQ-
M. hungatei      (654) QFLSQPWSNTEIISLMYTS-----------------------------------------
M. stadtmanae    (622) YYPSQSNIENINGREFTLGMDNGVIG----------------------------VSSP
M. jannaschii    (726) KREKGAFFKGTAYYLGNG-----------------------------------------T
M. maripaludis   (635) DREMGYYQQLGGSAQDING-----------------------------------TTVVYI
M. voltae        (689) EREAGSFQGLQTYATNIND-----------------------------------TLIVRS
P. horikoshii    (613) DGSPFPYVVYLTPSLGVLAYYKVATSNFVKLAFGIPTSSYSEFAEKLFSNFIPVYQYGSV
P. abyssi        (614) DGSPFPYVLYLAPNFGLITYYKVATSNFIKLAFNIPISKYSGFTEKLYSNFVPVYGYGNV
P. furiosus      (563) RSGVYVYVN--------------------------------------------LDYGYA
T. kodakarensis  (581) ESGAYLYIN--------------------------------------------LNYNYA
A. fulgidus      (541) -----------------------------------------------------------
D. vulgaris      (578) ISSIPQQLSYSLDTGEVQVQG----------------------------------NTTGV
S. solfataricus  (611) KAYNSLIVQMFIESLYQIG-----------------------------------------
C. jejuni        (538) TFSTAYPLDVKNGEIYLSN----------------------------------GVV
S. cerevisiae    (595) -----------------------------------------------------------

H. marismortui   (726) PSERVSAMEHYRLVGSS-------------------------------------------
H. volcanii      (799) PEERVSALEHYRLVKSSNSSALRSG----------------------------------
H. walsbyi       (809) PTHRVEALDQYRLVHVNNQSAR-------------------------------------
N. pharaonis     (723) PGERTEALEHFRLVHAGG-----------------------------------------
M. hungatei      (710) PITDTPALKNYRLIYESNG----------------------------------------
M. stadtmanae    (677) TNETNGSTMTFAYVDQSK-----------------------------------------
M. jannaschii    (781) ILANVNVYTYSYVTLINS-----------------------------------------
M. maripaludis   (690) PLQETDSYRVINILEITD-----------------------------------------
M. voltae        (744) LIQQTAEYNIYTLIEVRN-----------------------------------------
P. horikoshii    (666) IVYEFRPFAIYKIEDFINGTWREVGKLSPGKHTLRLYISAFGRDIKNATLYVYALNGTKI
P. abyssi        (668) IVYEFRPFAIYRIEELINGTWKAVNSLTPGKHELKLYISAFGRDIRNATLYVYAIG--NK
P. furiosus      (612) ILANEKAWETNLLRLFTQ-------------------------------------------
T. kodakarensis  (633) ILMNGEAFNTNLARLFIK-------------------------------------------
A. fulgidus      (592) -----------------------------------------------------------
D. vulgaris      (633) LASTIDVFDEGQLERRDY-----------------------------------------
S. solfataricus  (666) YGQPIAPFTTQIVPTSSG-----------------------------------------
C. jejuni        (593) LSDDFRSFKIGDNVVSVN-----------------------------------------
S. cerevisiae    (650) -----------------------------------------------------------
```

```
H. marismortui  (757) ------------ETEAFPNNPRYQSS---AWAKIFERVPGATIEGTG-PANATVRAAVQ
H. volcanii     (828) -----SYQRSLISEGNTYGLQPQALVPNNPAWVKTFERVPGATVDGSGAPANTTVTARVQ
H. walsbyi      (838) -----SILRTTYRGAQAAGINPRTTVPPQPSYVKTFERVPGATIEGSNAPTNSTVTAQVE
N. pharaonis    (753) -----------PVQPSMFELLMTQQVGAGDQWVKTFERVDGATIEGTG-PAESEIQAAVE
M. hungatei     (729) --------------------TQTFPGDVTLNNIKIFERVKGYTIPG--------------
M. stadtmanae   (707) -------------------------LNKSLNMSTLEDKQMRSKELSDGTGN--------
M. jannaschii   (800) -------------------------TNISTAIVQKINGQAKIIGT--------------
M. maripaludis  (715) -------------------------SEIKSANAVIDSNNQTSMQS--------------
M. voltae       (769) -------------------------ETLTGAMMAVTNDGQMQTQQ--------------
P. horikoshii   (726) IKRIKVGEIKYMNHLEEYPIIVNVTLPTAQKYRFILAQKGPVGVLTGPVRVNGKITNPAY
P. abyssi       (728) TEKIKIGEIEYMNHLNEKPIIVNVTLPKAEKYRLVLVQKGPVGVLTGPPKLNGEIANPIR
P. furiosus     (627) -------------------------RTGENYELVYSDGGFVKFVR--------------
T. kodakarensis (648) -------------------------PDGP-YELVYSDGGIIKILR--------------
A. fulgidus     (592) ------------------------------------------------------------
D. vulgaris     (659) ------------------------------------------------------------
S. solfataricus (685) ------------------------------------------------------------
C. jejuni       (615) ------------------------------------------------------------
S. cerevisiae   (650) ------------------------------------------------------------

H. marismortui  (774) MRNPATNKTFVYRQRTQTDQNGNFEMTVPYSTTGYDDWGTDEGYTNVSVRAETQYQFTAI
H. volcanii     (853) MRDLTTGTNFTYTQQAQTDADGEFTMTLPYSTTGYDEYGPDNGYTNVSVRAAGGYAFTGP
H. walsbyi      (861) LTIPNKNSTFTYTQQAQTTDTGAFSMTVPYSTTEYDEYGPDNGYTDISVHANDSYTISSA
N. pharaonis    (771) IDDG-TGEAFTYTQFAETDADGNFEMTVPYATTGYDEYGVEEGYTNTDVRANTTYQFIEM
M. hungatei     (748) ----------------------------TGTIEVPIVTNQGRHFTYRQQSENGTFTLP--
M. stadtmanae   (725) -------------------------TLLKPHKLIVVENNQLTEKIVNNNSNMSIMAIHQN
M. jannaschii   (818) -------------------------FKIHKLYIKTPLG----VKELVLNKDGQLSEFIRI
M. maripaludis  (733) -------------------------PNFHKLILKVNGN----VYEQETNENGDYSEIVRL
M. voltae       (787) -------------------------LNMHKVKLMVNENGKSKMYNSLADPDGQLSLLIKV
P. horikoshii   (786) IMREGESGRLELKVGVDKEYTADLYLRATFIYLVRKGGKSNEDYDASFEPHMDTFFITKL
P. abyssi       (786) IAREGEKGTLSLKVGVDKDYTADLYLRATFIYLVRKEGKSNDDYNAAFEPHMDTFFITKL
P. furiosus     (645) -------------------------FVHPNVVFR---GNKFILTGNGTG--LGLYGYLDN
T. kodakarensis (666) -------------------------LKHPNVVVERRGGETLLQFENATGTRLGIWGFLDN
A. fulgidus     (592) ------------------------------------------------------------
D. vulgaris     (677) -------------------------IVRQGGSGNAIADYAARQRNIESRRNIHFLFNKV
S. solfataricus (703) ----------------------------------YTITGSPLPRVQLMYFQPAYIALFPV
C. jejuni       (633) ------------------------------SIVEINSIKQGEYKITPIDDKAQFYIFYLK
S. cerevisiae   (650) --------------------------VRQQMITPLDVPPLDYFDEVFTSENWMVRIYQL

H. marismortui  (877) GANNGNRTGFTGATE------VTEGQVIGEDDDAATVELEPVTIQQ---GNGTSGES---
H. volcanii     (968) TSVTGNSTIVSYQAEN---VAVDEGLVNGAEDGTVQVTLERNEQELDLPGDSSSEDSSSE
H. walsbyi      (976) PSFNESGYVVSYQAN----VSVPEGLVNGDESGAIDVELDRNAQELSIGGSSSDSDSGSD
N. pharaonis    (878) EEG------MIAEAD------VSEAQVIGEDDSPVAVEMEP-FEESDLTDPAPDGDVTAP
M. hungatei     (804) ----------------------YATTNSPYDVKTTGPYRIIETNQTIDVDESQIEKYYT--
M. stadtmanae   (786) DGS--------------YFTVLFDSHLEESLFTKLYLKSGLNVTRFNMTHSEPGISVWDV
M. jannaschii   (869) EA------DG------RGYAWLATRNLEDSIYAKLHFLDGYGLKHIKLVKATIDPTDFGI
M. maripaludis  (784) EK----LSDG----TYQVYAWVSSKNLEDSIYTKLHFLDGYGLEKISLEKESVDPTSYGI
M. voltae       (842) DKNSIIGTDGSNNPVYSSSSWMATANLEDSVYSKLHFFDGEGLDTIKLEKESLDPTANGV
P. horikoshii   (906) KEGIKLRPGE--------NEIVVNAEMPKNAISSYKEKLEKEHGDKLIIRGIRVEPVFIV
P. abyssi       (906) KGGIKLHKGD--------NVVTAELNMPNGVISSYKEKLEKEYGDKLIIRGIRVEPVFIA
P. furiosus     (695) GTLVFKKWYS--------VKNMQEFELPNNLNGSVVVRYVYTQKIVLDRGIVRVKN----
T. kodakarensis (720) GTKVFEKWYN--------VEGLKEFELPAEVNG-TVIRYAYAEGEKILDRGVFRRD----
A. fulgidus     (592) ------------------------------------------------------------
D. vulgaris     (711) TG---------------ELKVLDDRLYNTVMVQLLLCSFDDTRFAPYFKLIYDNVYARV
S. solfataricus (729) GNG------------------GAGGLYYTVYIMVMVYQFTQPGTVLKPTVMVNS-----
C. jejuni       (663) DSAIP-----------YAQFILMDKTMFNSAYVQMFFLGNYDKNLFDLVINSRDAKVFKL
S. cerevisiae   (683) KK--------------------DDAQGRTLRDVGELTRSSTKTRRSIKRPELGLRV---
```

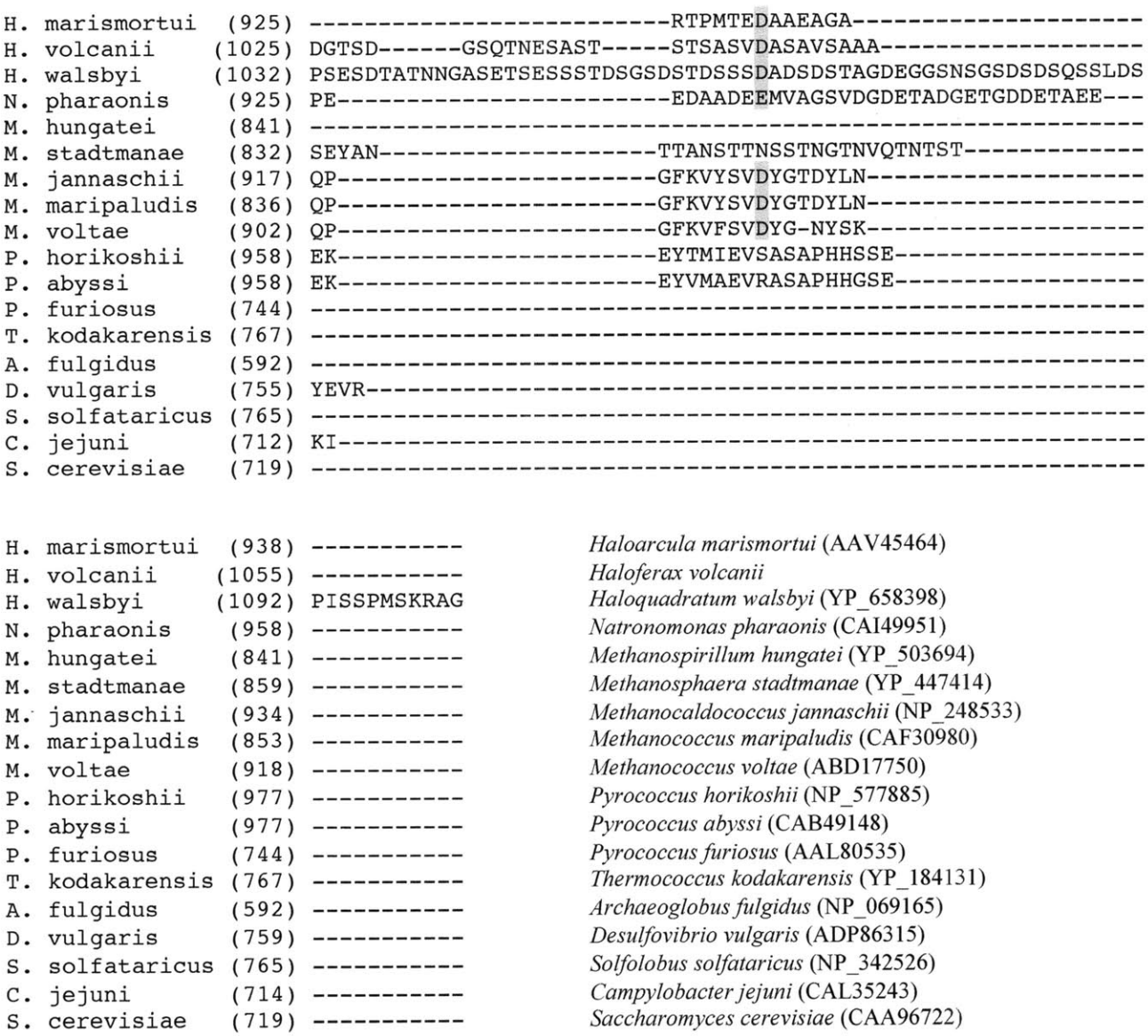
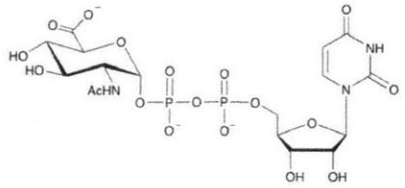
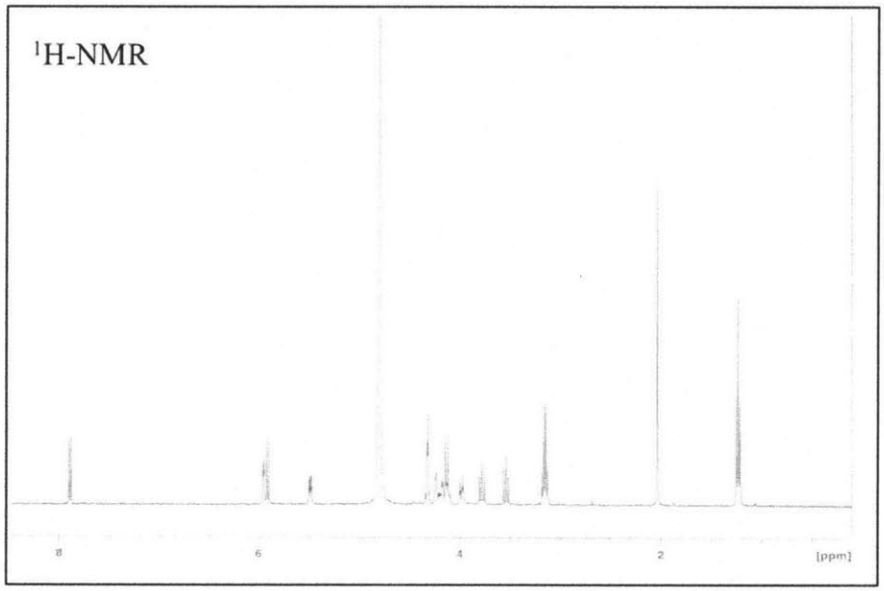


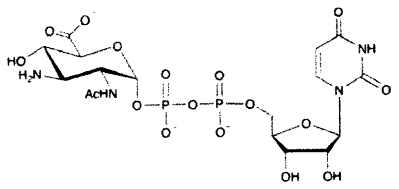
Figure 1: Sequence alignment of archaeal OTases with PglB (*C. jejuni*) and Stt3 (*S. cerevisiae*). Homologs were obtained using BLAST, and the alignment was performed with ClustalW. Conserved residues are highlighted in black and similar residues are indicated in gray. The GenBank accession number of each homolog is listed.

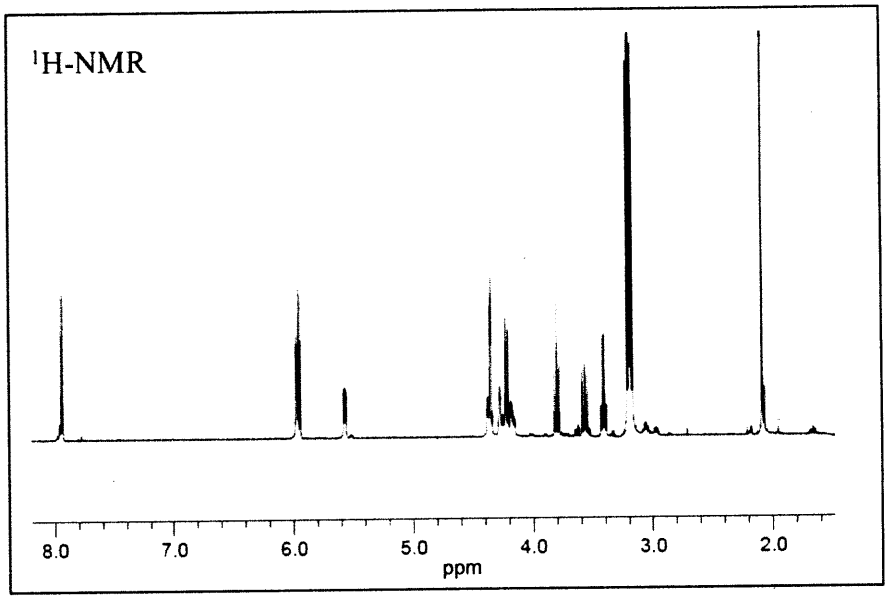
UDP-N-acetyl-D-glucosaminuronic acid (UDP-GlcNAcA)

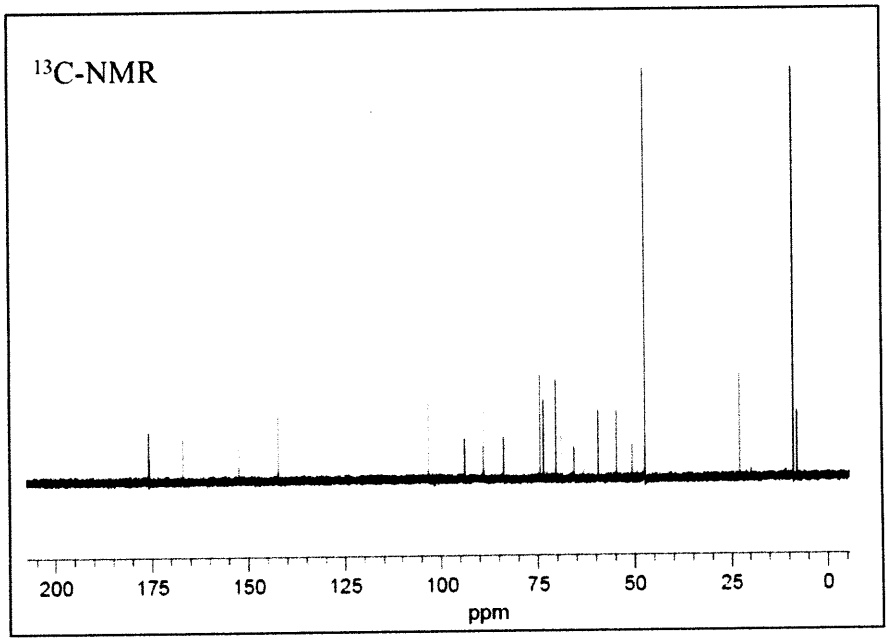




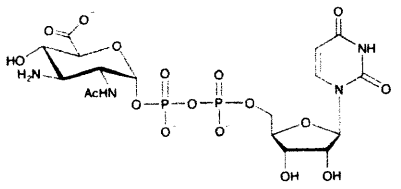
UDP-2-acetamido-3-amino-2,3-dideoxy-D-glucuronic acid (UDP-GlcNAc(3NH$_2$)A)

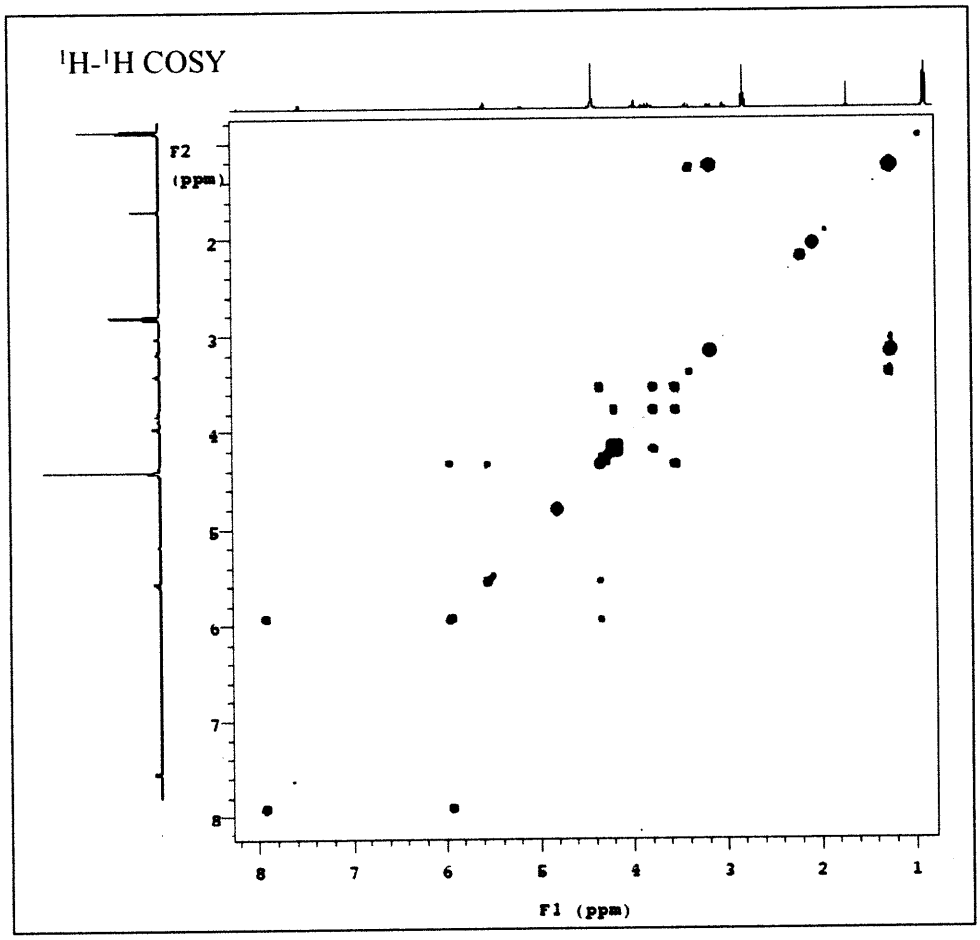




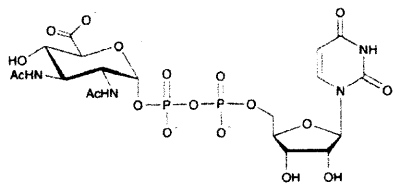


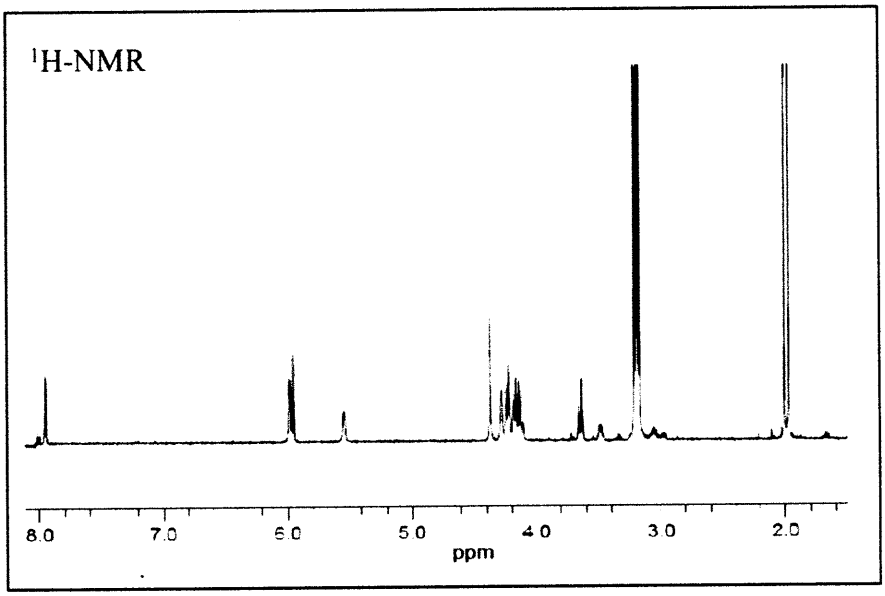
UDP-2-acetamido-3-amino-2,3-dideoxy-D-glucuronic acid (UDP-GlcNAc(3NH$_2$)A)

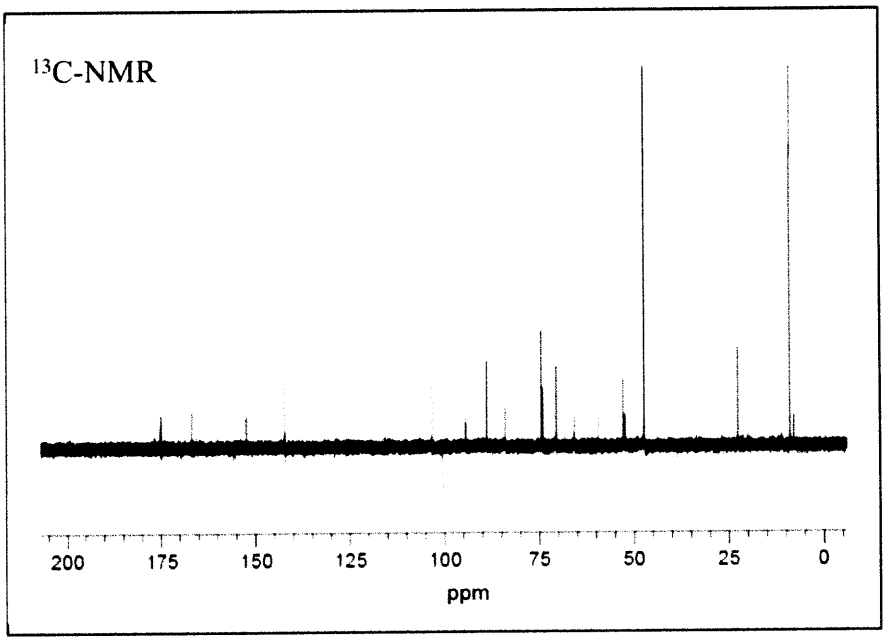




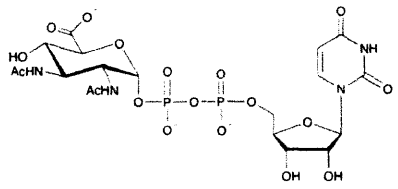
UDP-2,3-diacetamido-2,3-dideoxy-D-glucuronic acid (UDP-GlcNAc(3NAc)A)

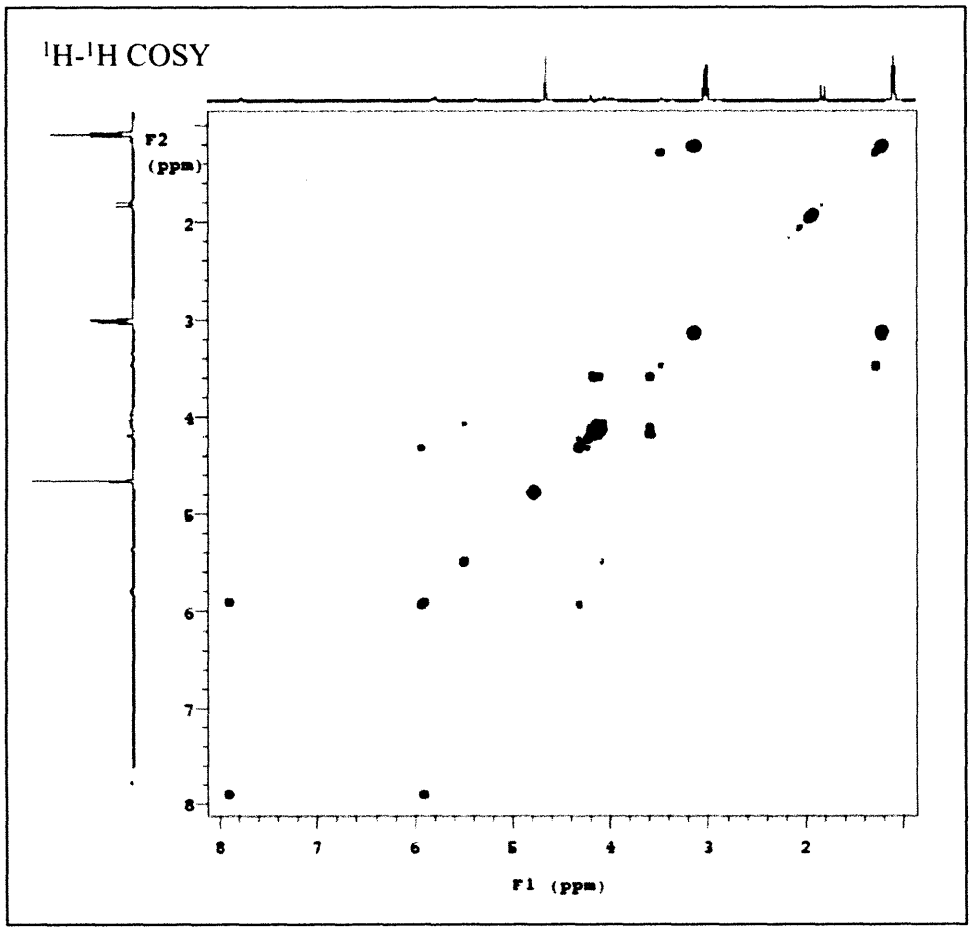




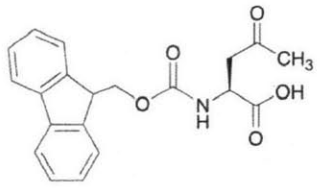


UDP-2,3-diacetamido-2,3-dideoxy-D-glucuronic acid (UDP-GlcNAc(3NAc)A)

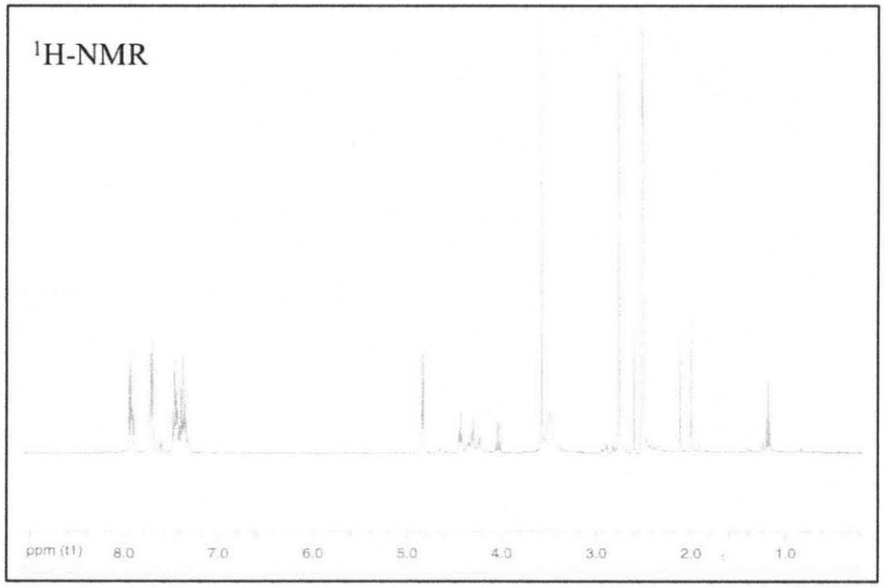




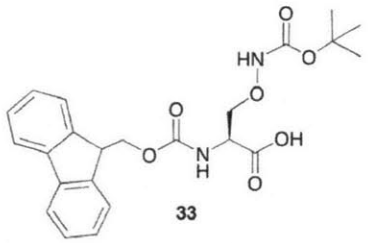
(L)-Fmoc-Oxonorvaline-OH



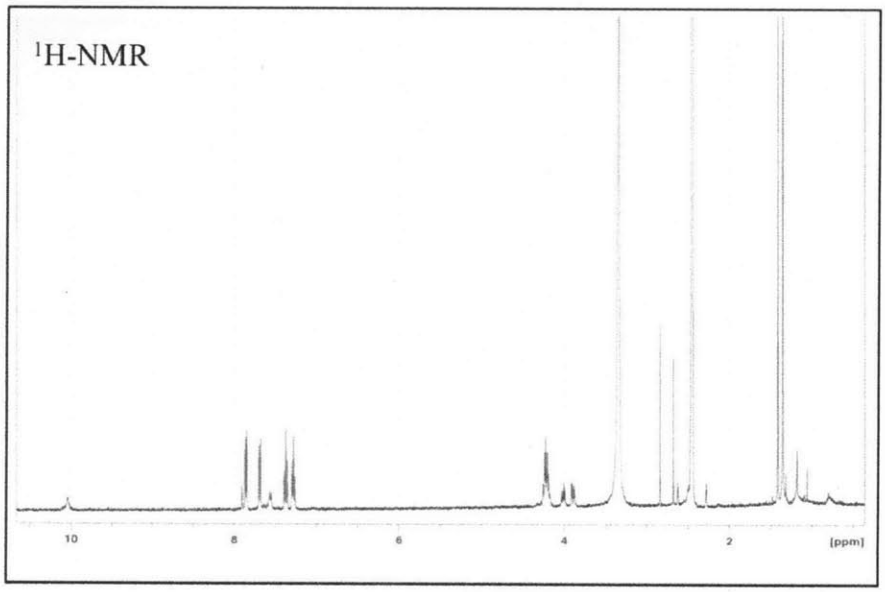
29



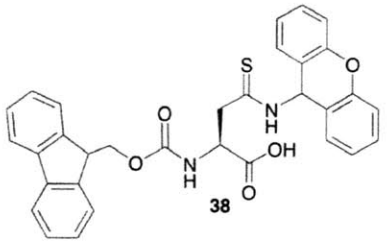
(L)-Fmoc-Aminoserine-(Boc)-OH

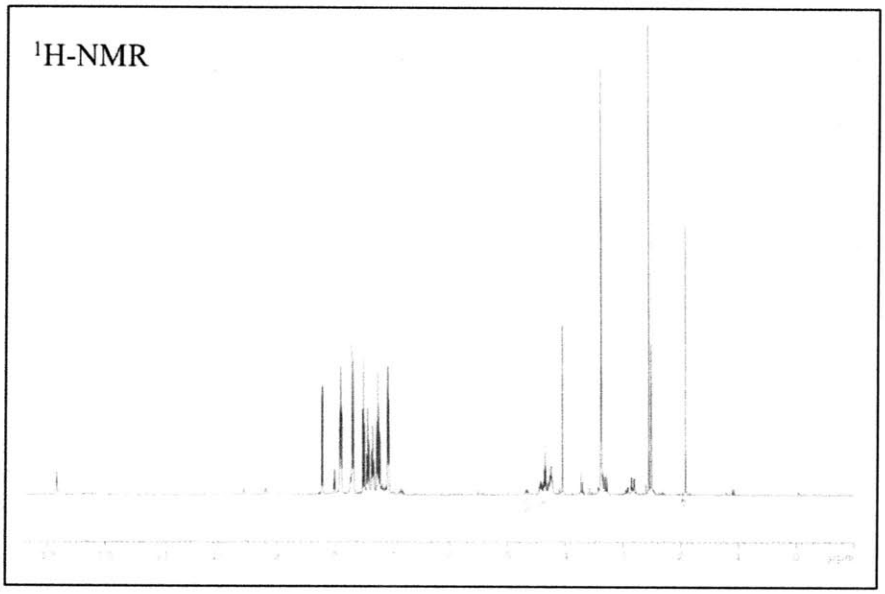


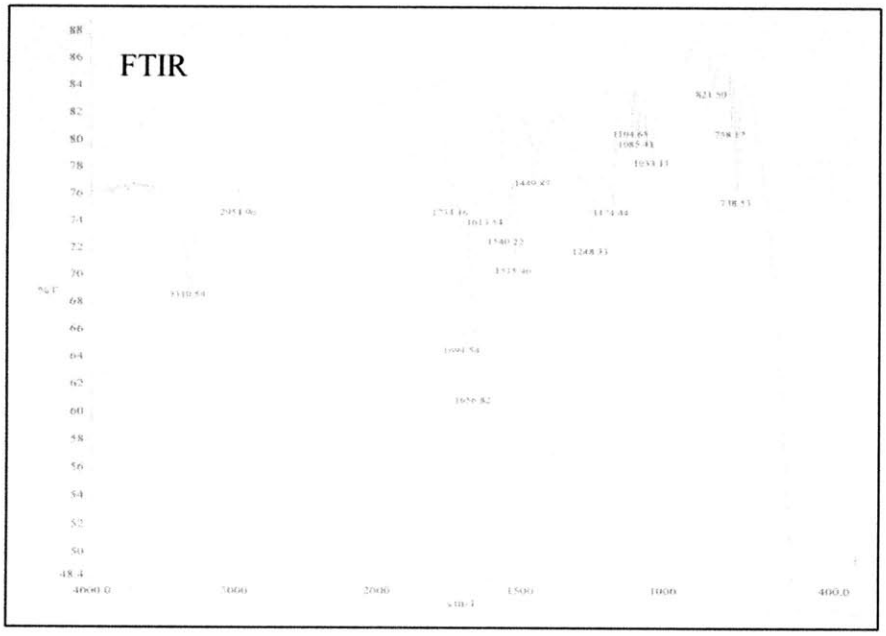
33



(L)-Fmoc-Thioasparagine-(xanthen-9-yl)-OH







Angelyn Larkin

Department of Chemistry
Massachusetts Institute of Technology
77 Massachusetts Avenue, 18-563
Cambridge, MA 02139
(617) 452-2826; alarkin@mit.edu

Education:

2005-2010 **Massachusetts Institute of Technology**, *Cambridge, MA*
Ph.D., Organic Chemistry

1997-2002 **Wellesley College**, *Wellesley, MA*
B.A., Chemistry and Political Science with Honors

Research Experience:

2005-2010 **Massachusetts Institute of Technology**, *Cambridge, MA*
Graduate Research, Advisor: Professor Barbara Imperiali
Investigation of asparagine-linked glycosylation in prokaryotes:
- Identification of a target prokaryotic oligosaccharyl transferase from *M. voltae* for in depth study using molecular biology and protein expression
- Biochemical characterization of enzymes in the Wbp pathway of *P. aeruginosa* for chemoenzymatic synthesis of the rare nucleotide sugar, UDP-GlcNAc(3NAc)A
- Discovery and validation of a novel NAD^+-recycling pathway in the dehydrogenase/aminotransferase pair WbpB and WbpE of *P. aeruginosa*
- Solved the X-ray structure of WbpE, an aminotransferase from *P. aeruginosa*, bound to its nucleotide sugar product and PLP cofactor.
- Synthesized a panel of asparagine isosteric peptides as potential inhibitors of the bacterial oligosaccharyl transferase, PglB

2003-2005 **Seattle Genetics, Inc.**, *Bothell, WA*
Research Associate, Department of Chemistry
- Development of antibody-drug conjugates as targeted cancer therapeutics

1997-2002 **Wellesley College**, *Wellesley, MA*
Undergraduate Research, Advisor: Professor David Haines
- Synthesis of adenosine analogs based on 1,2,4-triazoles

Teaching Experience:

Spring 2008:	Teaching Assistant, Chemical Tools for Assessing Biological Function (Graduate)
Spring 2006:	Teaching Assistant, Organic Chemistry I (Undergraduate)
Fall 2005:	Teaching Assistant, Principles of Chemical Science (Undergraduate)
2003-2005:	University Tutoring Service, Science/Math Tutor for College Students
Summer 1999:	English Teacher, Shanghai High School, Shanghai, China

Awards and Affiliations:

May 2009:	Wyeth Scholar, Department of Chemistry, MIT
March 2009:	Janette Houk Fellowship, Department of Chemistry, MIT
December 2002:	Honors in Chemistry, Department of Chemistry, Wellesley College
May 1998:	First Year with Distinction, Wellesley College

Sigma Xi Honor Society, member

American Chemical Society, member

Publications:

Larkin, A.; Olivier, N.B.; Imperiali, B. Structural Analysis of WbpE from *Pseudomonas aeruginosa* PAO1: A Nucleotide Sugar Aminotransferase Involved in O-Antigen Assembly, **2010**, *Biochemistry*, 49, 7227-7337.

Larkin, A; Imperiali, B. Biosynthesis of UDP-GlcNAc(3NAc)A by WbpB, WbpE, and WbpD: Enzymes in the Wbp Pathway Responsible for O-antigen Assembly in *Pseudomonas aeruginosa* PAO1, **2009**, *Biochemistry 48*, 5446-55.

Hartley, M.D; Larkin, A; Imperiali, B. Chemoenzymatic Synthesis of Polyprenol Phosphates. *Bioorg. Med. Chem.* **2008**, *16,* 5149-56.

Invited Talks:

"Harnessing Biosynthetic Enzymes from the Wbp Pathway of *Pseudomonas aeruginosa* for the Study of Asparagine-Linked Glycosylation," Gordon Conference on Carbohydrates, Glycobiology session, Tilton, NH, June 2009

Poster Presentations:

Larkin, A; Imperiali B. "Harnessing Biosynthetic Enzymes from *Pseudomonas aeruginosa* for the Study of Asparagine-Linked Glycosylation," American Chemical Society National Meeting, Boston, MA, August 2010.

Larkin, A; Imperiali B. "Harnessing Biosynthetic Enzymes from the Wbp Pathway of *Pseudomonas aeruginosa* for the Study of Asparagine-Linked Glycosylation," Alberta Ingenuity Centre for Carbohydrate Science, National Carbohydrate Symposium, Banff, AB, Canada, May 2010.

Larkin, A.; Imperiali, B. "Harnessing Biosynthetic Enzymes from the Wbp Pathway of *Pseudomonas aeruginosa* for the Study of Asparagine-Linked Glycosylation," Gordon Conference on Carbohydrates, Tilton, NH, June 2009.

Hartley, M.D.; **Larkin, A.**; Imperiali, B. "Utilization of an undecaprenol kinase for investigations of oligosaccharide transfer," American Chemical Society National Meeting, Boston, MA, August 2007

Meyer, D.L.; Burke, P.J.; Larkin, A.; Zabinski, R.F.; Moser, R.F.; Lewis, T.; Leia, S,; Stone, I.; Benjamin, D.R. "An Antibody-FRET conjugate that sheds light on intracellular antibody-drug conjugate activation," American Association of Cancer Research Annual Meeting, 2007.

Li, Y; Li, Hongli, L; Wang, M.N.; Bassi, R.; Mangalampalli, V.; Ludwig, D.; Balderes, P.; Witte, L.; Zhu, Z.; Meyer, D.; Larkin, A.; Senter, P.; Bohlen, P.; Hicklin, D. "Suppression of Leukemia Growth In Vivo by a Novel Anti-FLT3 Antibody-Auristatin Conjugate," American Association of Cancer Research Annual Meeting, 2005.

Li, Y; Li, Hongli, L; Wang, M.N.; Bassi, R.; Ludwig, D.; Witte, L.; Zhu, Z.; Bohlen, P.; Meyer, D.; Larkin, A.; Senter, P.; Hicklin, D. "Suppression of FLT3-Expressing Leukemia by a Monoclonal Antibody-Auristatin Conjugate," American Society of Hematology National Meeting, 2004.

Larkin, A.K.; Haines, D.R. "Synthesis of selectively [C-3] and [N-4]-substituted 1,2,4-triazoles," American Chemical Society National Meeting, New Orleans, LA, March 2003.

Haines, D.R.; **Larkin**, **A.K.**; Doran, A.C.; Wittenhagen, L.M. "Synthesis of selectively ribosidated 1,2,4-triazoles," American Chemical Society Meeting, San Francisco, CA, March 2000.